

# Traffic Simulation using Web Information of Activities Location\*

Songhang Chen and Fenghua Zhu, *Member, IEEE*

**Abstract**—Information of Activities location is critical for traffic simulation based on activity. To replace time-consuming manual acquisition, this paper proposes an information extraction method based on ontology and frequent subtree to acquire place-related information from the Web automatically. A destination selection model based on Nested Logit is established to use this information in traffic simulation. To verify the method and model, information of places within Zhongguancun area is acquired from the Web, and then experiments with uniform design are carried out on artificial transportation systems and regression analysis is done.

## I. INTRODUCTION

Many activity-based traffic simulation systems (ATSS) have been developed, such as TRANSIM, VISEM, FAMOS, TransModeler and TransWorld [1]-[5]. It is widely accepted that, comparing to traditional traffic simulation, ATSS possess many advantages, such as the capability to study people's travel actions in a deep and fine-grained way [6]-[8]. ATSS have become one hot topic in transportation modeling and analysis for a long time. However, they still face many difficulties, among which the most notable one is how to acquire activity information. For most existing ATSS, the acquisition and configuration of activity information is accomplished manually. As involved information is huge, much time and energy is wasted inevitably.

On the other hand, World Wide Web, which is the world's largest open source information platform, has been involved into people's daily life and has a profound effect on the development of human society. Information on the Web is all-inclusive and a great number of people create or share information and knowledge online. These characteristics have attracted researchers' keen attention. Sakaki *et al.* uses Twitter as a sensor for the real-time sensing of messages about earthquake [9]. Steinberger *et al.* present a system to monitor disease epidemics by text mining from the Web [10]. Zeng *et al.* show applications of web information in emergency response [11]. Asur and Huberman demonstrate how social media can be used to predict box-office revenue for movies [12]. In the field of transportation, Carvalho *et al.* [13] propose a text classifier based on Support Vector Machines (SVM) for

real-time sensing of traffic information in massive Twitter messages. Bregman *et al.* [14] gives a survey of uses of social media in public transportation. Collins *et al.* [15] use social media and conduct sentiment analysis to evaluate transit riders' satisfaction. Pereira *et al.* [16] utilize the web as a predictor for public transport demand around event venues. Internet is also used to improve traditional traffic survey [17] [18]. However, up to now, to our knowledge, there is still a lack of specific methods for conducting activity-based traffic simulation using information acquisition from the Web.

This paper aims to explore activity-based traffic simulation using activity-related information acquisition, especially place-related information acquisition from the Web. The main contributions of this paper include three folds:

- 1) *Propose an information extraction method to obtain place-related information from the Web automatically;*
- 2) *Set up a destination selection model using the place-related information for activity-based traffic simulation;*
- 3) *Design and carry out computational experiments based on artificial transportation systems to verify the method.*

The rest of paper is organized as follows. The method to obtain place-related information from the Web is described in Section II. In Section III, a destination selection model is proposed to utilize this web information. Validation based on artificial transportation systems is carried out in Section IV. Finally, in Section V, conclusions are drawn and future works are discussed.

## II. PLACE-RELATED INFORMATION ACQUISITION FROM THE WEB

An information extraction method for activity places is introduced in this section in three steps. First, the formal information model of activity places is established. Second, the information extraction process is presented. Finally, a solution for a challenge in the process is discussed.

### A. Ontology-based Place Model

Before extracting web information, the specification of place-related information must be defined. An ontology, which is a *formal, explicit specification of a shared conceptualization* [19], can provide a common semantic foundation for different subjects and has been adopted widely in information extraction. Here, we also use it to set the formal information model of places. OWL [20] is the most popular language to construct ontology now. An ontology with OWL can be defined as a tetrad  $\Omega := \langle C, I, P, X \rangle$ .  $C$  represents *owl: class* which defines a group of individuals that share some properties.  $I$  denotes collection of individuals, which are instances of classes.  $P$  is the collection of properties

\*This work was supported in part by National Natural Science Foundation of China projects 60921061, 70890084, 90920305, 90924302, 60904057, 60974095, 61174172, 61101220, 61104054 and 61004090; CAS projects 2F09N05, 2F09N06, 2F10E08, 2F11D03 and 2F11D01.

Songhang Chen is a Ph.D. student at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences (e-mail: chensohg@gmail.com).

Fenghua Zhu is an associate research scientist with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, 100080, China.

Songhang Chen and Fenghua Zhu are also with Dongguan Research Institute of CASIA, Cloud Computing Center, Chinese Academy of Sciences, Songshan Lake, Dongguan, 523808, China.

(*rdf:Property*) used to state relationships between individuals or from individuals to data values.  $X$  denotes the axiom set used to associate class and property identifiers. Classes can be organized in a specialization hierarchy using *owl:subClassOf*. A built-in class named *owl:Thing* is the class of all individuals and a superclass of all OWL classes.

Fig.1 shows partial place ontology constructed with OWL in Protégé [21]. *Place* class, *Road* class (which *place* relies on) and *Vacation* class (which *place* belongs to) are three top classes. According to types of destination of people's daily activities, five subclasses, *ResidentialCommunity*, *School*, *Hospital*, *BusinessPlace*, and *OfficeBuilding*, are derived from *place*. Further, *BusinessPlace* is classified into *BodybuildingPlace*, *EntertainmentPlace*, *Restaurant* and *Shop*. *Place* class has general properties of a place, like ID, name, address, longitude and latitude. As well as these properties, its subclasses have their own properties. For example, *BusinessPlace* class has properties for service time, and its subclass *Restaurant* has properties like average consumption, style of cooking and some information with characteristics of the Web (e.g. number of netizens' comments and their grades for taste, environment and service of restaurants).

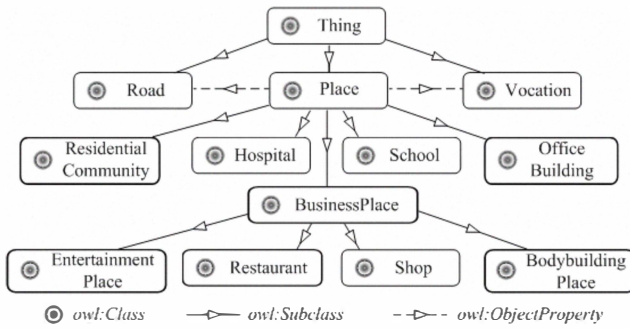


Figure 1. Place ontology.

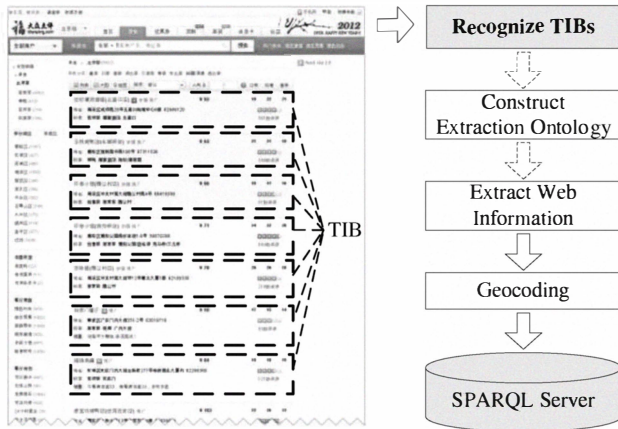


Figure 2. Target Information Blocks (TIBs) and the extraction process.

## B. Process of Web Information Extraction

With the development of e-commerce, many review or life-guide sites appear on the Internet, such as dianping.com, soufun.com and yelp.com. They offer a lot of information about places for people to retrieve. Moreover, retrieved pages usually present results in a list layout, as Fig.2 shows.

Information of *place* individuals is distributed in independent information blocks, called as *Target Information Blocks* (TIBs) in this paper. Since most retrieved pages are assembled automatically according to databases, their layouts are almost the same, especially HTML structures of TIBs. Thus, we propose a method based on the place ontology to extract place information. The whole process consists of five steps.

### Step 1: Recognize TIBs

Besides TIBs, retrieved pages often contain much useless content, such as navigation links, advertisement column, JavaScript codes and copyright information. How to recognize TIBs automatically and accurately is critical for the process. A method based on frequent subtree is implemented and will be described in detail in *Part C* of this section.

### Step 2: Construct extraction ontology

In this step, the mapping relations between information items (contained in TIB) and the place ontology are calibrated, i.e., map information items to classes and properties. For the convenience of management and maintenance, this mapping information is also stored in an ontology with OWL. It must be noted that this step can be done offline easily.

### Step 3: Extract web information

In the previous step, positions of information items in TIB are recorded. Using this position information, open-source tools such as *HTML Parser* and *JTidy* can be employed to extract these items from mass TIBs automatically.

### Step 4: Geocoding

In this step, address text is transferred into geographical coordinates. Spatial coordinate is necessary for a place to be located in a road network, but it rarely appears in common web pages. Here, the free geocoding service offered by Google is used to transfer address text (e.g. Haidian Street A.36) into coordinates (e.g. latitude 39.9825008 and longitude 116.3094487) [22].

### Step 5: Storage of instances

In this step, extracted information is assembled into *place* classes. A semantic database, SPARQL server is adopted as the storage system. It supports SPARQL language [23] and provides great convenience for information management by allowing database operations in a semantic way.

## C. Recognition of TIBs based on Frequent Subtree

As mentioned before, recognition of TIBs is the first and the most important step in the whole process of information extraction. Therefore, we will discuss it in detail.

Instead of the entire Web, we prefer retrieving information from some famous vertical websites, which collect and organize a mass of information by categories and are usually built well. According to W3C standards, a retrieved page can be expressed as a DOM (Document Object Model) tree, every node on which corresponds to a tag in HTML, as Fig.3 shows.

**Definition 1:** A *rooted ordered tree* [24] is a directed acyclic graph satisfying (1) there is a distinguished vertex called the root that has no entering edges, (2) every other vertex has exactly one entering edge, (3) there is a unique path from the root to every other vertex, and (4) there is a predefined

ordering among each set of siblings. The order is implied by the left-to-right order in figures illustrating an ordered tree.

**Definition 2:** For a *rooted ordered tree*  $T$  (either ordered or unordered) with vertex set  $V$  and edge set  $E$ , we say that a rooted tree  $T'$  (either ordered or unordered, depending on  $T$ ) with vertex set  $V'$  and edge set  $E'$  is a *bottom-up subtree* [25] of  $T$  if and only if (1)  $V' \subseteq V$ , (2)  $E' \subseteq E$ , (3) the labeling of  $V'$  and  $E'$  is preserved in  $T'$ , (4) for a vertex  $v \in V'$ , if  $v \in V'$  then all descendants of  $v$  (if any) must be in  $V'$ , and (5) if  $T$  is ordered, then the left-to-right ordering among the siblings in  $T$  should be preserved in  $T'$ . Intuitively, a bottom-up subtree  $T'$  (with the root  $v$ ) of  $T$  can be obtained by taking a vertex  $v$  of  $T$  together with all  $v$ 's descendants and the corresponding edges.

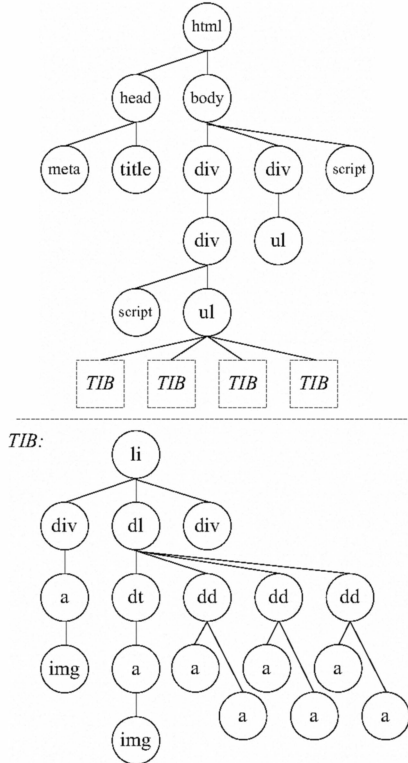


Figure 3. The DOM tree of a retrieved web page.

Clearly, utilizing the **definition 1** and **2**, the DOM tree can be treated as a *rooted ordered tree* and every TIB can be treated as its *bottom-up subtree*. Thus, the problem to recognize TIBs is transferred to search *frequent bottom-up subtrees*. Then, with the common path from the root to the subtrees, we can extract the HTML of TIB from web pages.

To facilitate the searching process, preorder strings are adopted to express DOM trees. A preorder string  $S$  of tree  $T$  is defined recursively as follows:

$$S = \begin{cases} \gamma 0 \\ \gamma S_1 \dots S_n 0 \quad n \geq 1 \end{cases} \quad (1)$$

where  $\gamma$  is the label of root of  $T$ ,  $S_1 \dots S_n$  is the preorder strings of each subtree under  $\gamma$  from left to right, and 0 is a special

symbol to indicate the end of a preorder string. It has been proved that preorder strings contain the same number of label and 0 [26]. Therefore, we can get preorder strings of all bottom-up subtrees of tree  $T$ : scan from any HTML label  $\ell$  in the preorder string  $S$  till the number of label and 0 passed is the same, and the scanned substring is just the preorder string of the bottom-up subtree with  $\ell$  as the root.

#### D. Algorithm

The algorithm is implemented with Java. The inputs consist of a web document for extraction and a group of restrictions: the minimum length  $\psi$  of a TIB, the minimum similarity  $\theta$  and the minimum supports  $\Omega$ . The output is the common path of TIBs. Firstly, the algorithm calls the API of *JTidy* to build the DOM tree  $T$  of the web document and gets its preorder string  $S_r$  by a preorder traversal. Secondly, scan  $S_r$  from left to right to acquire preorder strings of all bottom-up subtrees of  $T$  and filter out those with a length shorter than the minimum length  $\psi$ . Thirdly, sort these preorder strings in the lexical order for ease of later comparison. Lastly, calculate the similarity of adjacent preorder strings based on edit distance [27]. If the similarity is greater than the minimum similarity  $\theta$ , treat them as the same and count their frequency. If the frequency is greater than the minimum supports  $\Omega$ , output the path from the root to the corresponding bottom-up subtree as a result. In practice, choosing proper restrictions ( $\psi$ ,  $\theta$  and  $\Omega$ ) can make the algorithm return just one path.

### III. DESTINATION SELECTION MODEL BASED ON NESTED LOGIT

After acquiring placed-related information from the Web, we further explore how to integrate the retrieved place-related information into activity-based traffic simulation.

In traffic simulation, a destination selection model is used to decide the destination of travel and often implemented with discrete choice models (DCM) [28] [29]. Among various DCM, Multinomial Logit Model (MNL) and Nested Logit Model (NL) are the two of the most popular. Especially, NL is based on MNL but overcome its IIA (Independence of Irrelevant Alternative) problem [30].

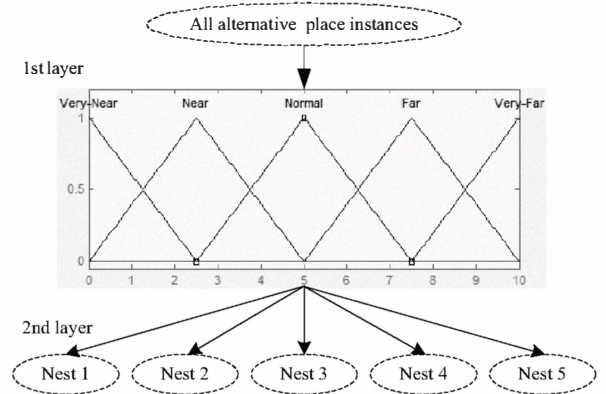


Figure 4. Destination selection model based on Nested Logit. The horizontal axis of the membership function is distance in kilometers.



A new destination selection model based on NL is with a layered structure (see Fig.4). The first layer is the distance factor. Five fuzzy linguistic terms (Very-Near, Near, Normal, Far, and Very-Far) are used to express the distance perception and judged by triangular membership functions. Through the first layer, all alternative places will be divided into five nests, as the second layer shows. Any two nests are disjointed, and it is possible that some nests are empty.

If a person is at place  $i$  and the type of place for the next activity is  $k$ , then the probability  $\Pr(j)$  for the place  $j$  (belongs to  $k$ -type) to be chosen can be calculated as follows (For detailed derivation of a general two-layer NL, please see [31]).

$$\Pr(j) = P_{j|m} \cdot P_m = \frac{(e^{V_j})^{1/\theta_m}}{\sum_{l \in N_m} (e^{V_l})^{1/\theta_m}} \cdot \frac{\left( \sum_{l \in N_m} (e^{V_l})^{1/\theta_m} \right)^{\theta_m}}{\sum_{k \in M} \left( \sum_{l \in N_k} (e^{V_l})^{1/\theta_k} \right)^{\theta_k}} \quad (2)$$

$$V_j = \alpha_k \cdot d_{ij} + \bar{\beta}_k^T \cdot \bar{x}_k \quad (3)$$

where  $V_j$  is a deterministic component of place  $j$ 's utility calculated from observed variables;  $d_{ij}$  is the distance from place  $i$  to place  $j$ ;  $\alpha_k$  is the distance coefficient of the  $k$ -type place;  $\bar{x}_k$  is a vector of observed attributes of the  $k$ -type place;  $\bar{\beta}_k$  is a vector of coefficients corresponding to  $\bar{x}_k$ ;  $N_m$  is the set of places in the nest  $m$ ;  $P_m$  is the probability for the nest  $m$  to be chosen;  $P_{j|m}$  is the probability for the place  $j \in N_m$  to be chosen if the nest  $m$  has been chosen; and  $\theta_m$  ( $\theta_m \in [0,1]$ ) is the correlation coefficient of alternatives' utility in the nest  $m$ . When  $\theta_m = 0$ , random components of alternatives' utility in the nest  $m$  are totally dependent; when  $\theta_m = 1$ , they are totally independent and meanwhile NL is reduced as MNL.

Using the choice of restaurant as an example, the retrieved restaurant information contains grades for taste, environment and service, denoted as  $x_1$ ,  $x_2$  and  $x_3$  respectively. Let  $\bar{x} = [x_1, x_2, x_3]^T$ ,  $\bar{\beta} = [\beta_1, \beta_2, \beta_3]^T$  and  $\theta_m = 0.5$  ( $m = 1, 2, \dots, 5$ ). If a person is at place  $i$  and its next activity is eating out, then the probability for the restaurant  $j \in N_m$  to be chosen is:

$$\Pr(j) = P_{j|m} \cdot P_m = \frac{(e^{V_j})^2}{\sum_{l \in N_m} (e^{V_l})^2} \cdot \frac{\sqrt{\sum_{l \in N_m} (e^{V_l})^2}}{\sum_{k=1}^5 \sqrt{\sum_{l \in N_k} (e^{V_l})^2}} \quad (4)$$

$$V_j = \alpha \cdot d_{ij} + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 \quad (5)$$

#### IV. VALIDATION BY EXPERIMENTS ON ARTIFICIAL TRANSPORTATION SYSTEMS

Computational experiments are carried out on Artificial Transportation Systems (ATS) to verify the effectiveness of our method. ATS [32]-[34] is an extension of traditional transportation systems from the complex systems perspective.

It adopts agent-based modeling and simulation methods to model each participant in the real traffic world as one agent. By simulating each agent's activities under the *simple-is-consistent* principle, complex traffic phenomena can emerge from bottom to up [35]. Based on this idea, a microscopic transportation simulation platform named TransWorld has been developed [36]. On this platform, the works of this paper are verified in the following steps: 1) prepare related data for modeling and simulation and especially use the proposed method to acquire place-related information from the Web; 2) reconstruct TransWorld to read place data from the SPARQL server instead of the previous database; 3) implement the new destination selection model in TransWorld.

##### A. Data Preparation

Zhongguancun area, which is a prosperous business district in Haidian District, Beijing, is selected as transportation simulation area. The modeled road network is shown in Fig.5, which is a screenshot of an ATS-modeling system based on Google Map [37]. With the operation panel, we can draw directly main roads and intersections of the Zhongguancun area on the web map. Information hid in the background such as geographical coordinates, intersection channelization and connective relationship can be configured simultaneously.



Figure 5. Road network modeled manually in an ATS-modeling system.

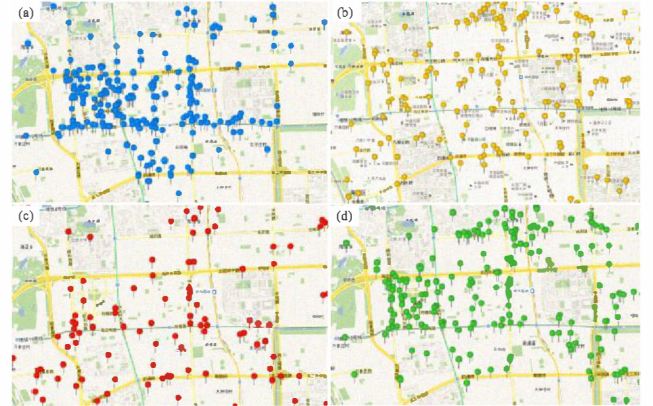


Figure 6. (a) ~ (d) tags of restaurants, bodybuilding places, residential areas and entertainment places within Zhongguancun Area on the web map respectively.

After analyzing the structure of life-guide websites, and their classification of places and geographic division, we use a web crawler to download web pages. These pages contain information about places within Zhongguancun area. The total

amount of download pages achieves 1.3 GB. Using the information extraction method proposed in Section II and some data cleansing technologies, we finally get information of 8163 restaurants, 1587 residential communities, 2010 entertainment places, 5993 shops, 592 bodybuilding places, 386 office buildings and 2779 places for other services (e.g. hotel, hospital and school). Fig.6 (a)-(d) illustrates partial restaurants, bodybuilding places, residential areas and entertainment places respectively. Every tag on the web map represents a place. Due to space limitations, detail of data cleansing is beyond this paper.

### B. Experiment Design

Travel simulation of TransWorld involves many models which contain different variables and coefficients. Here, we just focus on the destination selection model and take the choice of restaurants as an example. Table I shows the value range of each coefficient in the formulae (5). To investigate impacts of variables (distance, taste, environment and service), uniform design method [38] [39] is used to find out a group of optimal coefficients.

TABLE I. RANGE OF PARAMETERS

	$\alpha$	$\beta_1$	$\beta_2$	$\beta_3$
Range	[-5,0]	[0,5]	[0,5]	[0,5]

Suppose that the experimental domain consists of  $s$  factors  $x_1, \dots, x_s$ . Without loss of generality, this domain can be assumed as a unit cube  $C^s = [0,1]^s$ . The aim of the uniform design is to choose a set of  $n$  experimental points  $\rho = \{x_1, \dots, x_n\} \subset C^s$  that are uniformly scattered on  $C^s$  (measured with discrepancy  $D$ ) to represent all experimental points. A uniform design table is denoted by  $U_n(n^s)$ , where  $U$  denotes the uniform design,  $n$  is the number of experiments and  $s$  is the number of factors. Here,  $U_{10}(10^4)$  (see Table II) is adopted. Every factor is divided into ten levels, concrete values of which are in brackets. In the same case, 10,000 experiments and 100 experiments are needed in full factorial design and orthogonal design respectively. However, using uniform design, only 10 experiments are needed, and great efforts are saved.

TABLE II. UNIFORM DESIGN TABLE  $U_{10}(10^4)$

No.	$\alpha$	$\beta_1$	$\beta_2$	$\beta_3$
1	1(-0.5)	4(2.0)	3(1.5)	5(2.5)
2	5(-2.5)	6(3.0)	7(3.5)	1(0.5)
3	8(-4.0)	2(1.0)	2(1.0)	2(1.0)
4	4(-2.0)	7(3.5)	1(0.5)	9(4.5)
5	6(-3.0)	3(1.5)	10(5.0)	6(3.0)
6	7(-3.5)	10(5.0)	4(2.0)	7(3.5)
7	10(-5.0)	8(4.0)	5(2.5)	4(2.0)
8	9(-4.5)	5(2.5)	8(4.0)	10(5.0)
9	2(-1.0)	9(4.5)	9(4.5)	3(1.5)
10	3(-1.5)	1(0.5)	6(3.0)	8(4.0)

### C. Experiment Result

TransWorld can record detail information of all agents' eating-out activity, including departure time, arrival time, travel distance and probability of restaurants when they are chosen as destinations. Therefore, average travel time of agents' eating-out activity can be calculated after every experiment and treated as the optimization target. With the coefficients in Table II and keeping the other coefficients fixed, we conduct ten groups of experiments, and every group is repeated five times to reduce stochastic errors. Statistical analysis of every group experiment is shown in Fig.7. The small black dots are data points, and the column denotes the data mean. The bars show 95% confidence interval of the mean.

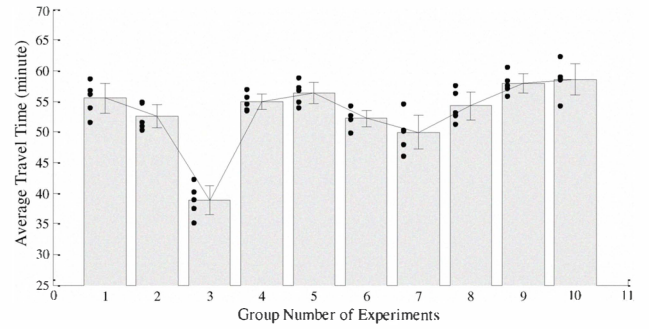


Figure 7. Average travel time of eating-out activity.

It can be observed that the Group 3, with distance coefficient  $\alpha=-4.0$ , taste coefficient  $\beta_1=1.0$ , environment coefficient  $\beta_2=1.0$  and service coefficient  $\beta_3=1.0$ , gets the minimum travel time. We randomly select 20% samples of this group for regression analysis with a significant level of 0.5. The multiple linear regression equation is denoted as:

$$y = b_0 + b_1d + b_2x_1 + b_3x_2 + b_4x_3 \quad (5)$$

where  $b_0, b_1, b_2, b_3$  and  $b_4$  are regression coefficients;  $d$  is travel distance;  $x_1$ ,  $x_2$  and  $x_3$  are grades for taste, environment and service of a restaurant respectively;  $y$  is the probability of a restaurant to be chosen. The result of stepwise regression indicates that every argument is significant and their contributions to the regression equation can be sorted as  $x_3 > d > x_1 > x_2$ . The service and distance factors are in the first two places to affect the probability for a restaurant to be chosen. This is consistent with our usual experience. To verify or calibrate the model further, actual traffic survey is needed, which is a part of our future work.

### V. CONCLUSION

This paper proposes an information extraction method based on ontology and frequent subtree to acquire place-related information from the Web for activity-based travel simulation. Practice proves that it is effective to gain a lot of place information automatically online and can replace previous manual information acquisition and configuration. The destination selection model based on Nested Logit model is adopted to fuse this web information in an easy and general

way. Experiments and regression analysis of the results show that the model is rational to some extent.

#### ACKNOWLEDGMENT

We thank Professor Fei-Yue Wang for his intensive instructions. We also thank the Lab of Complex Adaptive Systems for Transportation (CAST), Prof. Shuming Tang, and all hard works of Yisheng Lv and Peijun Ye. Without their works and supports, we cannot finish this work.

#### REFERENCES

- [1] Smith, L.; Beckman, R.; Baggerly, K.; etc., "TRANSIMS: Transportation analysis and simulation system", Proceedings of the 5th National Conference on Transportation. Planning Methods Application, Transportation Research Board, Washington, D.C, pp.1-10, 1995.
- [2] M. Fellendorf; T. Haupt; U. Heidl; W. Scherr, "VISEM - An Activity Chain Based Traffic Demand Model", Paper presented at the conference' Activity based approaches: Activity scheduling and the analysis of activity patterns, Eindhoven University of Technology, 1995.
- [3] Pendyala, R M, "Florida Department of Transportation, Phased Implementation of a Multimodal Activity-based Travel Demand Modeling System in Florida", Research report, Prepared for the Florida Department of Transportation, 2004.
- [4] Lu Lu; Tianlong Yun; Li Li; Yuelong Su; Danya Yao; "A Comparison of Phase Transitions Produced by PARAMICS, TransModeler, and VISSIM", IEEE Intelligent Transportation Systems Magazine, vol.2, no.3, pp.19-24, 2010.
- [5] Zhang Nan; Wang Fei-Yue; Zhu Fenghua; Zhao Dongbin; Tang Shuming, "DynaCAS: Computational Experiments and Decision Support for ITS", IEEE Intelligent Systems, vol.23, no.6, pp.19-23, 2008.
- [6] C. R. Bhat; F. S. Koppelman, "Activity-based Modeling of Travel Demand", Springer, 1999.
- [7] W. Davidson; R. Donnelly; Vovsha, Peter; etc., "Synthesis of first practices and operational research approaches in activity-based travel demand modeling", Transportation Research Part A: Policy and Practice, vol. 41, pp. 464-488, 2007.
- [8] R. Kitamura; S. Fujii, "Two computational process models of activity-travel behavior", Theoretical foundations of travel choice modeling, pp. 251-279, 1998.
- [9] T.Sakaki; M.Okazaki; Y.Matsuo, "Earthquake shakes twitter users:real-time event detection by social sensors," Proceedings of the 19th international conference on World wide web. NewYork, NY, USA:ACM, pp.851-860, 2010.
- [10] R. Steinberger; F. Fuat; E. Van der Groot; C. Best; P. Etter; R. Yangarber, "Text Mining from the Web for Medical Intelligence," in Mining Massive Data Sets for Security, OIS Press, 2008. Netherlands.
- [11] Zeng Da-Jun; Wang Fei-Yue; Cao Zhi-Dong, "Open Source Information in Emergency Response", Science & Technology Review, vol.16, no.6, pp.27-33, 2008.
- [12] S.Asur; B.A.Huberman, "Predicting the future with social media," Mar2010. [Online]. Available: <http://arxiv.org/abs/1003.5699>.
- [13] Sara Carvalho; Lu's Sarmento; Rosaldo J. F. Rossetti, "Real-Time Sensing of Traffic Information in Twitter Messages", Proceedings of the IEEE ITSC Workshop on Artificial Transportation Systems and Simulation, 2010.
- [14] Bregman, Susan, "Overcoming Barriers to Using Social Media in Public Transportation", TRB Annual Meeting, Washington, D.C, 2012.
- [15] Collins, Craig; Hasan, Samiul; Ukkusuri, Satish V, "Novel Transit Rider Satisfaction Metric: Riders' Sentiments Measured from Online Social Media Data", TRB Annual Meeting, Washington, D.C, 2012.
- [16] Pereira, Francisco C.; Rodrigues, Filipe; Ben-Akiva, Moshe E, "Internet as Sensor: Case Study with Special Events", TRB Annual Meeting, Washington, D.C, 2012.
- [17] Kalenoja, Hanna Katariina; Tiikkaja, Hanne Pilvi Pauliina; Rantala, Jarkko Veli-Matti, "Role of Internet Surveys in Travel Diary Surveys", TRB Annual Meeting, Washington, D.C, 2012.
- [18] Diana, Marco, "Internet Mobility Survey Sampling Biases in Measuring Frequency of Use of Different Transport Modes", TRB Annual Meeting, Washington, D.C, 2012.
- [19] Studer R; Benjamins V R; Fensel D., "Knowledge Engineering, Principles and Methods". Data and Knowledge Engineering, vol.25, no.122, pp.161-197, 1998.
- [20] Wang, X.H.; Zhang, D.Q.; GU, T.; Pung, H.K, "Ontology based context modeling and reasoning using OWL", Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications, pp. 18- 22, 2004.
- [21] Gennari, John H.; Musen, Mark A., Fergerson, Ray W, etc., "The evolution of Protégé: an environment for knowledge-based systems development", International Journal of Human-Computer Studies, vol.58, no.1, pp.89-123, 2003.
- [22] Google Geocoding API [EB/OL]. <https://developers.google.com/maps/documentation/geocoding/?hl=en>.
- [23] Malik, S.; Goel, A.; Maniktala., "S. A comparative study of various variants of SPARQL in semantic web", IEEE International Conference on Computer Information Systems and Industrial Management Applications (CISIM), pp.471-474, 2010.
- [24] Yun Chi; Richard R. Muntz; Siegfried Nijssen; Joost N. Kok., "Frequent Subtree Mining - An Overview", Fundam. Inf, vol.66, no.1-2, pp.161-198, 2004.
- [25] Valiente, G., "Algorithms on Trees and Graphs", Springer, 2002.
- [26] Luccio, F.; Enriquez, A. M.; Rieumont, P. O.; Pagli, L., "Exact Rooted Subtree Matching in Sublinear Time", Technical Report TR-01-14, Università Di Pisa, 2001.
- [27] Marzal, A.; Vidal, E., "Computation of normalized edit distance and applications," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.15, no.9, pp.926-932, Sep 1993.
- [28] M. Ben-Akiva; S. Lerman, "Discrete Choice Analysis: Theory and Application to Travel Demand", MIT Press, Cambridge, MA, 1985.
- [29] Train K, "Discrete Choice Methods with Simulation". Cambridge: Cambridge University Press, 2003.
- [30] McFadden, D., "Modeling the choice of residential location", Transportation Research Record 672, TRB, National Research Council, Washington, DC, pp.72-77, 1977.
- [31] Chieh-Hua Wen; Frank S, "Koppelman. The generalized nested logit model", Transportation Research Part B, 35, pp.627-641, 2001.
- [32] Wang Fei-Yue; Tang Shuming, "Concepts and Frameworks of Artificial Transportation Systems", Complex Systems and Complexity Science, vol.1, no.2, pp.52-59, 2004.
- [33] Wang Fei-Yue, Tang Shuming, "Artificial Societies for Integrated and Sustainable Development of Metropolitan Systems", IEEE Intelligent Systems, vol.19, no.4, pp. 82-87, 2004.
- [34] Wang Fei-Yue; Tang Shuming, "A Framework for Artificial Transportation Systems: From Computer Simulations to Computational Experiments", Proceedings of 8th IEEE International Conference on Intelligent Transportation Systems, Vienna, Austria, pp.1130-1134, 2005.
- [35] Tang Shuming, Wang Fei-Yue, Liu Xiaoming, et al. "A Preliminary Investigation on Artificial Transportation Systems", Journal of System Simulation, vol.17, no.3, pp.704-709, 2005.
- [36] Wang Fei-Yue, "Parallel Control and Management for Intelligent Transportation Systems: Concepts, Architectures, and Applications", IEEE Transactions on Intelligent Transportation Systems, vol.11, no.3, pp.630-638, 2010.
- [37] Pejic, A.; Pletl, S.; Pejic, B., "An expert system for tourists using Google Maps API," IEEE 7th International Symposium on Intelligent Systems and Informatics, pp.317-322, Sept. 2009.
- [38] Fang K.T, "Uniform Design and Uniform Design Table", Beijing, 1994.
- [39] Fang, K.T; Wang, Y, "Number-theoretic methods in Statistics", Chapman and Hall, London, 1994.