

Ontology based Suggestion distribution system

Tong Lee Chung, Bin Xu, Xuanyu Yao, Qi Li, and Bozhi Yuan

Department of Computer Science and Technology, Tsinghua University, Beijing,
China

tongleechung86@gmail.com, xubin@tsinghua.edu.cn, yaouxuanyu1993@gmail.com,
lawby1229@163.com, zhongguoliqi@163.com

Abstract. The digitization of modern cities has brought cities to a new level. There are still many new areas yet to be discovered in this new ecosystem. Today, there is an urgent need for smarter cities to support the growing population. One particular problem is citizens do not know which city department to give their suggestions to. This paper presents a system for distributing suggestions from citizens to the right city officials based on ontology knowledge base. We use data from official websites to construct our ontology and do experiments with actual suggestions from citizens. The experiments show some promising results.

Keywords: Suggestion distribution, ontology application, city management

1 Introduction

City digitization has been going on in China for over a decade and it is believed to be the right time to begin focusing on smarter cities [1]. This is an exciting time of the century where new intelligent systems is expected to appear in cities. But, with great opportunities comes great challenges. Today, city management has become more sophisticated than ever before. Two main reasons can be concluded, one is the growing population and the other is the increasing complexity of city departments. One particular problem is for city officials to hear the voices to citizens and for citizens to give suggestions to the right officials. Today, the main strategy for dealing with is this is set up different portal for different departments or distributing suggestions manually. There are many practical issues in these two types of approaches. Citizens do not know the actual duties and roles of these departments, and it is not feasible to understand all duties and roles just to give a suggestion. With the huge city population, it is unreliable for humans to handle the tedious job.

Many researchs have been carried out for smart city development, including using semantic technology and ontologies. There have been many research with constructing ontology for city management and e-governance [2] [3] [4] and in the task of data integration using ontologies [5] [6]. There have also been a number of reseach on boosting the . But using these ontology is a challenging task. This paper investigates the possiblity of using ontology in city management and discuss some advantage of using ontology.

We envision that computers will be able to assist humans in doing many repetitive jobs in city management. We also believe that ontology technology will play an important role in future smart cities. In this paper, we present a suggestion distribution system based on ontology knowledge base. We design a system that takes a new suggestion as an input and gives the department that is most likely in charge of responding to the suggestion as an output. Our system uses an ontology to search for the corresponding department. Our method of searching for the corresponding department differs from many modern classification methods where the classification model is trained using another set of suggestions. Our approach is based on an ontology that describes the roles and duties of each department. Our system is flexible and robust to change. Our experiment shows that our system has satisfying results when compared to some baseline methods. We crawl data from official websites and implement a demonstration of the system.

The remaining of this paper is structured as follows. Section 2 will give a formal definition of our problem and challenges will also be discussed. In section 3, the details of implementation of the system will be presented. Section 4 will look into data for the experiment and experiment results. We will finally conclude our paper and point out directions for future research.

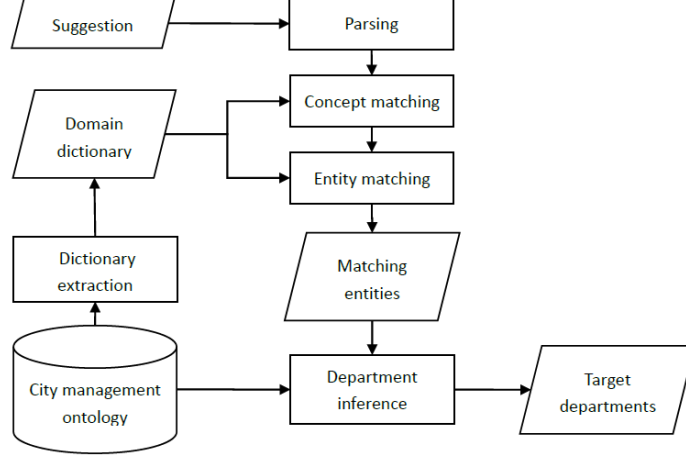
2 Implementation of Suggestion Distribution System

The process of inferring a department from a suggestion can be divided into two different parts. The first part is matching an entity that is most related to the suggestion. The second part is inferring the corresponding data from the ontology. Figure 1 shows the inferring process in detail. When a suggestion arrives, it is first matched to the most related concept in the ontology. This step is for pruning down the search space so that a trivial search is not necessary. The entities of the most related concept act as the candidate entities that will be used for matching in the next step. In the second part, reasoning in the ontology is done to infer the most corresponding department.

2.1 Concept Matching

Given a new suggestion, we would like to first know which concept it most probably belongs to. In an ontology, there are many concepts, and each concept has a number of entities. Using this information, we propose an unsupervised approach to match the suggestions. Our idea is to capture the similarity using TF-IDF [7]. We need to first create a feature vector that is used to calculate the similarity. Using the idea of term frequency inverse document frequency from information retrieval, we treat each entity in a concept as a document. To deal with the robustness of the entities, we also include all datatype labels and their values as part of their content. For example, the properties of a personnel concept include name, position, and age. These can be very useful when matching suggestions, for example when a suggestion contains a position, it is more likely a personnel

Fig. 1. Department inferring process of department from suggestion



concept. The matching weight of a word in a suggestion to a concept is calculated as:

$$\wp(w) = \log(tf_w + 1) \times \log(idf_w) \quad (1)$$

where term frequency for a word w in the suggestion is given by $f_w = f_w/|W|$, where f_w is the frequency of the word w , and $|W|$ is the total number of words in the suggestion. Inverse document frequency $idf_w = |N|/n_w$, where N is the total number of entities in a concept and N_w is the number of documents that contain the word w . So given the matching weight of every word in the suggestion, the matching weight of a whole suggestion is given as the summation of all its words matching weight:

$$\mathcal{R}(sugg) = \frac{1}{count(w)} \times \sum_{w \in sugg} \wp(w) \quad (2)$$

As every word has a non-negative weight that has a maximum of 1, the summation of every word will not exceed one. The tfidf weight is consider high when the word appears a lot in suggestion and in multiple documents. During the matching phrase, we choose the concept that has the highest matching weight, but in reality, a suggestion may belong to multiple concepts. In the selecting stage, we set a treatment for this by first selecting the concept with highest match and then choosing concepts that satisfies $weight > weight_{max}$.

2.2 Entity Matching

After getting the matching concepts, we consider matching only entities that belongs to the matched concept. To deal with robustness problem, we use the

same treatment as above and use entity label, datatype property, and value of datatype property as entity feature. We use cosine similarity to calculate the matching of the suggestion and the concept. Cosine similarity is calculated using:

$$cSim(sugg, concept) = \frac{sugg \times concept}{|sugg| \times |concept|} \quad (3)$$

where $|sugg|$ is a count of each word in the suggestion and $|concept|$ is the count of each word in the concept. $sugg \times concept$ is

2.3 Department reasoning

After finding the most corresponding entity, we need search for a related department. We assume that an entity is related to all its department when one of its office has a certain role and this entity has a direct or in-direct upward link to this role. By up-ward link, we consider the basic ontology, because of the structure of the ontology, a department has some office, office has some roles, these roles will include some processes, processes are made up of activity and these activity can include some personnel, objects, documents. This is the down flow of city management. By upward link, we mean that the links go from lower layer entities to department entity. To customize our own rule set in OWLIM-Lite rule set, we need to modify one of the pre-defined rule set and add our own rules.

And finally, we have to return a weight that determines the relatedness of a suggestion and a department. The first feature we consider is similarity weight of suggestion and matched entity. Another feature we consider is the length of the evidence, the longer the evidence the less relatedness of the department. The final weight of a department to a suggestion if the summation of all evidence that lead to the department. The relatedness is calculated as below:

$$rel_{dep}(sugg) = \frac{1}{R} \sum_r cSim(sugg, concept) \times lenght(r) \quad (4)$$

Where r is a route for a concept in matched concept set to its related department, R is the total number of routes to a department dep . $lenght(r)$ is the length of the route from concept to suggestion.

2.4 UI

As it is shown in the Figure 2, our ontology based suggestion distribution system provides a easy and convenient platform for users to access and deal with suggestion data. A list of suggestions will take place on the left of the screen, and detailed information, including title, content, user's advice, the suggested category provided by the system, etc., will be shown when one of the suggestions in the list is once clicked. If the user wish to let the suggestion distributed to a specific bureau, he can choose among the radio boxes and press "Confirm" button. The system can automatically apply different themes to suggestions, with the undistributed in red, the correctly distributed in green and the wrongly distributed in yellow, which provides convenience distinguish suggestions of different status.

Fig. 2. User Interface

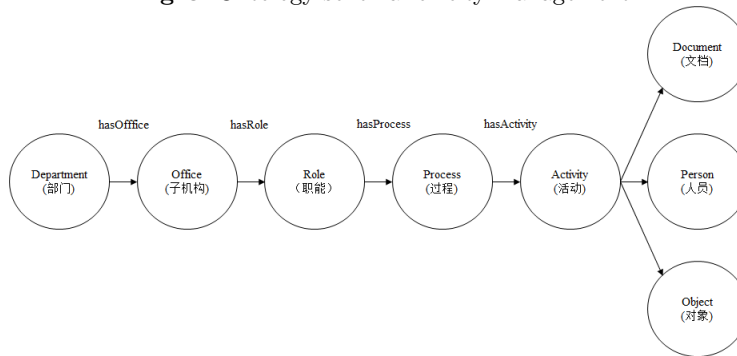


3 Experiment

3.1 Ontology construction

Constructing the city management ontology was a challenging task for the experiment. We first design a ontology schema for city management, Figure 3 shows the ontology schema for city management. We define eight classes in city management, namely, Department, Office, Role, Process, Activity, Document, Person and Object.

Fig. 3. Ontology schema for city management



We use web data to populate the city management ontology. We crawl city management guide from official websites and use it as source for constructing the city management ontology. We consider that the department and office are the corresponding department website and office. Then we populate the our ontology by using the management guide as process. Then by using document structure,

we can further populate our ontology. By using the document hierarchical structure, we capture management process knowledge of a certain department. Altogether, we pull 105 processes to construct the management ontology for these five departments.

3.2 Testing data

To get testing data to evaluate our system, we crawl the official website for suggestions. We consider the corresponding department as the target of the system. All of the suggestions that are posted on the web have a response. We clean our data using the response by removing the suggestion of which the response says ‘go to some other department’. After data cleaning, we choose 30 suggestions from each department to evaluate our system.

3.3 Results

We test our system using the test suggestion described in the previous section. We use two measures to evaluate our system. Top one match is when the department with the highest match is the target department. Since our system requires no training data, we use all data for testing. Another measure is top two match, where we consider that the system is correct when the target is in the two highest match. Table 1 shows the result of our system. Our system have a top one match of over 70 percent. From the results, we find that in a certain department the accuracy is very low. After looking at the suggestions that we crawled, we find that the suggestion isn’t relevant to any process in our ontology. The only way we can think of to fix this problem is to include related knowledge into our database.

Table 1. Experiment result for matching suggestions to department

	Correct	Total	Accuracy
Overall	66	90	73%
Traffic	16	30	53%
Civil Affairs	24	30	80%
Tourism	26	30	86%

We compare our method with some machine learning model with using knowledge from ontology. We choose Naive Bayes [8] classification method and Logistic Model [9] as our baseline method. We perform the a 5 fold cross-validation experiment using WEKA [10]. Table 2 shows that using ontology gives better performance compared to Naive Bayes and close to the performance of Logistic Model. Another important fact about our ontology based system is that it doesn’t depend on training data. Machine learning models will out-perform our system when training data large enough. Using ontology, on the other hand will have better performance if more knowledge is used.

Table 2. Experiment result for matching suggestions to department

	Accuracy
Ontology System	73%
NaiveBayes	66%
Logistic	75%

We believe that there is more potentials in our system, we can use better matching model in the system. But this paper shows that ontology system can performs just as well as machine learning models. The advantage of using ontology is that knowledge is highly portable and can be used in many different applications, whereas new data has to be required in most machine learning models. The ontology models the behaviour of city management and can be used solved related problem. This high reusability of data cannot be compared by most learning models.

4 Conclusion

With the rapid development of computer science technology, smart city applications can reach a certain level of intelligences. Semantic web technology is a set of tool that is especially suited for AI applications. This paper aims to discover some possible domains where semantic web technology can enhance city management. We look into the task of distributing suggestions of citizens to a corresponding city department. We propose a ontology based suggestion distribution system. Our system uses a city management ontology to refer entities that is most relevant to the suggestion and infers the department using reasoning tools. The main contribution of this paper is to prove the possibility of using ontology to assist humans in city management. We design a framework for implementing semantic technology in city management systems. Finally, we test our system using real life data from the web. The experiments show that our system have some promising results. The advantage of ontology is that the knowledge is portable and can be used in other systems and applications, using the same framework, we can implement other similar systems.

For our future work, we are looking into two different directions. One is to come up with more ideas that can enhance city management. The other direction is methods to populate the city management ontology using web data.

Acknowledgement: This work is supported by the China National High-Tech Project (863) under grant No SS2013AA010307

References

1. Kehua Su, Jie Li, and Hongbo Fu. Smart city and the applications. In *Electronics, Communications and Control (ICECC), 2011 International Conference on*, pages 1028–1031, Sept 2011.

2. S.K. Goudos, N. Loutas, V. Peristeras, and K. Tarabanis. Public administration domain ontology for a semantic web services e-government framework. In *Services Computing, 2007. SCC 2007. IEEE International Conference on*, pages 270–277, July 2007.
3. LeonidasG. Anthopoulos and Athena Vakali. Urban planning and smart cities: Interrelations and reciprocities. In *The Future Internet*, volume 7281 of *Lecture Notes in Computer Science*, pages 178–189. 2012.
4. John Fraser, Nick Adams, Ann Macintosh, Andy McKay-Hubbard, TomsPariente Lobo, PabloFernandez Pardo, RafaelCaadas Martnez, and JessSobrado Vallecillo. Knowledge management applied to e-government services: The use of an ontology. In MariaA. Wimmer, editor, *Knowledge Management in Electronic Government*, volume 2645 of *Lecture Notes in Computer Science*, pages 116–126. Springer Berlin Heidelberg, 2003.
5. Jun Zhai, Jiatao Jiang, Yi Yu, and Jianfeng Li. Ontology-based integrated information platform for digital city. In *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on*, pages 1–4, Oct 2008.
6. Nilesh Anand, Mengchang Yang, J.H.R. van Duin, and Lori Tavasszy. Genclon: An ontology for city logistics. *Expert Systems with Applications*, 39(15):11944 – 11960, 2012.
7. Thomas Roelleke and Jun Wang. Tf-idf uncovered: A study of theories and probabilities. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '08*, pages 435–442, New York, NY, USA, 2008. ACM.
8. George H. John and Pat Langley. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, UAI'95*, pages 338–345, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.
9. Marc Sumner, Eibe Frank, and Mark Hall. Speeding up logistic model tree induction. In AlpioMrio Jorge, Lus Torgo, Pavel Brazdil, Rui Camacho, and Joo Gama, editors, *Knowledge Discovery in Databases: PKDD 2005*, volume 3721 of *Lecture Notes in Computer Science*, pages 675–683. Springer Berlin Heidelberg, 2005.
10. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009.