Research on the construction of conceptual hierarchical model based on air pollution

Xueying Zhai, Yong Sun* and Min Ji
College of Geodesy and Geomatics, Shandong University of Science and Technology, Qiangdao, Shandong Province, China

*Corresponding author: yong_sun@sdust.edu.cn

Abstract. In recent years, the problem of air pollution has become more and more serious, attracting close attention of mankind. The ontology system is constructed for the semantics of related words of air pollution to realize the sharing and standardization of semantics. Humans’ comments on the air conditions of the living environment are taken as corpus, and words related to air pollution are extracted from the corpus text to construct the ontology of air pollution. Divide the levels according to the attributes of air pollutants, and build and enrich the ontology system. This paper builds an ontology model based on air pollutants, divides hierarchical structure on the basis of concepts.

Keywords: air pollution, domain ontology, semantic analysis, ontology modeling

1. Introduction
With the informatization of the development of spatial information technology, people have acquired massive amounts of geographic data information. Practical selection, sharing and interoperability of these data have become a major problem currently facing. Humans have certain differences and flexibility in the cognition and expression of things, so strict formal descriptions and clear definitions are required in the semantics of geographic information [1]. "Ontology" was originally a philosophical concept, from the perspective of philosophy. Ontology is a systematic explanation or explanation of objective existence, and it is concerned with the abstract nature of objective reality [2]. Gruber [3] is the most famous and widely used definition of ontology: ontology is a clear and standardized description of conceptualization. Introducing ontology ideas into the subject field can restore existing data, knowledge and information from an object-oriented form to a reasonable semantic system so that computers can process and users can share [4-5]. As the research on geographic ontology, including natural language processing, knowledge engineering, knowledge representation, and many other artificial intelligence theories and technologies, has received the attention and research of many research teams, UCGIS has listed geographic ontology as one of the ten long-term research and challenges. At present, there are many researches on geographic ontology at home and abroad. In foreign countries, researchers from the University of Calgary in Canada use GeoCO ontology to represent geographic space and clustering domain knowledge, and solve the problem of not paying too much attention to domain knowledge and user goals [6]. Kuhn of Mushai University analyzes the object behavior in the text, extracts the car navigation domain ontology from the traffic information,
and conducts research and analysis on the object behavior in the text [7-8]. In China, Cao Cungen and others have conducted in-depth research on the formal ontology of knowledge in the key project "National Knowledge Infrastructure (CNKI)"[9]. Yi Rulan [10] et al. analyzed the semantic formalization of the concept of loess landforms, constructed the logical structure of the ontology, and established the ontological model of the loess landforms [11]. In recent years, with the rapid advancement of China’s urbanization and the continuous development of industrialization, air pollution has become the focus of environmental problems. Most areas of our country have suffered from air pollution, posing a huge threat to people’s travel and health. [12]. At present, in view of the severe current situation of air pollution, researchers from all walks of life pay attention to and study the problem of air pollution control in many aspects, but there is very little research on the field based on air pollution. Ontology can express the inherent characteristics of entities in fields such as the structure and relationship between concepts, that is, "shared conceptualization".

2. Data source
The main source channels of concept words can be divided into standardized word sources and free word sources. The source of standardized vocabulary mainly refers to the existing chemical vocabulary. The sources of free vocabulary are mainly comments on air quality pollution from various websites, such as Weibo and WeChat official accounts and related websites. The data sources in this article are mainly professional chemistry vocabulary, extracted from webpages of online questionnaires and complaint platforms. This article is mainly based on the air pollution source classification system of "Changzhutan Regional Air Pollutant Emissions", and fills in pollution words from the perspective of public perception. The sources of air pollutants mainly start from the following aspects: fossil fuel stationary combustion sources, process sources, mobile sources, solvent usage sources, agricultural sources, dust sources, biomass combustion sources, storage and transportation sources, waste gas treatment sources, and other emission sources. The state of air pollutants is divided into: particle pollutants mainly from two aspects: aerodynamic particle size (particulate matter) and pollutant form. The aerodynamic particle size is mainly based on the particle size, fly dust (particle size <=10um), dust fall (>10um), and visible particles (>100um). The forms of pollutants are divided into: smoke, fog, and dust. Gaseous pollutants: sulfur-containing compounds, nitrogen oxides, carbon oxides, sulfuric acid fumes, photochemical fumes, halogen compounds, amino acid gases, chlorine-containing compounds, organic odors, and gases. According to the nature of air pollutants, it is divided into primary pollutants and secondary pollutants.

2.1. Complaint data mining
In the age of advanced network technology, massive amounts of data will follow, and it is very necessary to maximize the mining of information on the network. The rapid development of economics and industrialization has caused the public's sensitivity to air pollution to surge. The public has higher requirements for the environment and air quality. Therefore, the public makes relevant speeches about the atmospheric environment they are in, and extracts air pollution words for a large amount of speech information. Table 1 is the comment information of Shandong Environmental Public Prosecution (letters and visits) platform. The categories of air pollution on the complaint platform are mainly divided into: dust, malodor or peculiar smell, oil fume, motor vehicle or mobile source, industrial waste gas, smoke and dust, and others.

In the Table 1 complaint text, words related to air pollution were extracted: black smoke, black smoke, pungent, coal-fired boiler, dust, odor, gas boiler, smoke, choking nose, choking eyes, sulfur smell, Strong pungent and peculiar smell.
Table 1. Pollution words extracted from the complaint text

| Complaint text                                                                                                                                                                                                 | Extract relevant pollution words |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|
| The brick kiln factory has black smoke billowing every day, no procedures, no environmental protection facilities, dusty, and you must wear a mask when you go out. There is a pungent smell nearby, which endangers the health of the people. The local environmental protection department failed to inspect it. | Black smoke, Billowing smoke, Pungent |
| A chemical plant in the west is secretly started these days, the smell is unpleasant, and the nose and eyes are choked.                                                                                           | Choking nose, Choking eyes         |
| The smell of sulfur wafting from the brick factory every night is particularly intoxicating.                                                                                                                   | Sulfur smell, Smoke               |
| In the second half of 2017, there was often a strong pungent odor at night, which was very inconvenient to travel. I personally smelled it rather disgusting. After drying clothes for two days outside, there would be a strange smell when I took it back to the dormitory. Here comes a bad influence! Hope to solve it! | Strong pungent, rare delicacy, nausea, Strange smell |

2.2. Internet survey
The rapid development of the Internet and the continuous expansion of the field and scale have provided tremendous convenience for online questionnaire surveys. Online questionnaire survey is a survey method based on the technological environment of the Internet that uses web page questionnaires, e-mail questionnaires, etc. to collect data. In this article, a network questionnaire survey is conducted on the atmospheric environment, and the region mainly focuses on Shandong Province. Using the web page questionnaire method, based on the characteristics of convenience, timeliness, flexibility and speed, obtain comment information on the intuitive feeling of air quality.

Table 2. Pollution words extracted from the text of the online questionnaire

| Web questionnaire text                                                                                                                                                                                          | Word mining |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| The weather in Huangdao is generally good, and the nearby factories sometimes emit some exhaust gas, which is a bit peculiar.                                                                                | Exhaust, Smell |
| Inhaling the air feels pungent than in the previous two years, and the haze often causes widespread respiratory diseases among students.                                                                      | smog, Pungent, Respiratory diseases |
| There is more haze during heating, and the red alert is reached in severe cases. There is more haze in winter and spring.                                                                                           | Heating, smog |
| Generally speaking, it is okay. There used to be sandstorms and dust, but now there is less groundwater (although it is very close to the Yellow River)                                                             | Sandstorm, dust |

3. Ontological level of air pollutants
There are many pollutants in the atmosphere, and there are many types of pollutants. Therefore, the overall classification of pollutants is of significant significance. In the "Code of Names of Air Pollutants" (HJ524-2009), the classification method of air pollutants is mentioned. According to the physical and chemical properties and structural characteristics of air pollutants, the line classification method is suitable for classification. The classification of air pollutants adopts a hybrid classification method based on line classification and supplemented by area classification. Follow the principle of science, according to the most stable attributes of air pollutants, the nature and structure of the pollutants and related indicators as the classification basis, combined with application habits. In this article, the classification and model construction of the atmospheric domain ontology are mainly from the following six aspects: the source of atmospheric pollutants, chemical composition and internal structure, pollutant forms, pollutant properties, pollutant toxicity and pollutant existence status.
3.1. According to the source of pollutants
Air pollutants are substances that are discharged into the atmosphere due to human activities or natural processes and have harmful effects on people and the environment. Air pollutants come from various sources, and are classified into natural pollution sources and man-made pollution sources according to the nature and characteristics of pollution sources. Natural pollution sources are divided into biological and non-biological. The organisms that pollute the atmosphere mainly include pathogens, molds, pollen, dust mites, etc. Non-biological manifestations are mainly volcanoes, earthquakes, waves, forests, fires, and methane decomposed by plants and animals. Man-made pollution sources are divided into production pollution sources and domestic pollution sources according to the process of pollutants. Common sources of production pollution are mainly industry, agriculture, and transportation the sources of life pollution are residential buildings, schools, hospitals, commerce, restaurants, public toilets, and animal husbandry.

3.2. According to the state of pollutants
Classified according to the state of air pollutants, it mainly includes aerosol pollutants and gaseous pollutants. Particulate pollutants are mainly classified according to the particle radius. The particle size is less than or equal to 10um is floating dust, mainly PM10, PM2.5, PM0.1, the common phenomenon is haze, fine powder; more than 10um and less than 100um are dust reduction, common Is the total suspended particulate matter (TSP). The particles larger than 100um are visible to the naked eye, and the common ones are white powder, plastic foam particles, dust particles, etc. The pollutants are mainly divided into: smoke, fog, and dust.

3.3. According to the nature of the pollutant
Gaseous pollutants discharged into the atmosphere from pollution sources can directly pollute the atmosphere, and at the same time, they can react to form secondary pollution. Among the secondary pollutants, photochemical smog is the most harmful and has received the most attention. The main types are: London smog, Los Angeles smog, industrial photochemical smog.
4. Ontology model construction

![Ontology hierarchy model]

The composition of the atmosphere is complex and diverse, and the terms describing air pollution are more diverse. Then it is particularly important to build a model for a large number of air pollution words to achieve a systematic arrangement. There are many attributes of air pollutants, mainly natural attributes, diffusibility, toxicity, activity and persistence. The meaning of attributes is that the monitoring data can be correctly judged and processed. In this article, the model is mainly constructed from multiple angles and different attributes. Mainly divided into 3 layers. The root node of the first layer is air pollution. The second layer structure is mainly based on the classification of pollution levels and attribute expansion. The third layer is the filling of the concept of pollution words.

5. Conclusion
The function of ontology has a wide range of applications, which can be used as a representation method of knowledge and applied to knowledge engineering and knowledge management. Ontology is faster and more convenient for storing knowledge. In the field of air pollution, there are few researches on ontology concepts. This article mainly conducts text mining from massive texts. Comprehensively classify and sort out the concept of pollution words in the field of air pollution. A more comprehensive understanding of the knowledge system based on the field of air pollution. Ontologies can also be used to express and store knowledge.

Acknowledgments
This work was financially supported by Shandong Province Major Scientific and Technological Innovation Project "Provincial Natural Resources Monitoring and Supervision Big Data Application Service Platform Construction (2019JZZY020103); Shandong Province Key R&D Project (No.2016GSF117017) fund.

References
[1] Shang SY, Zhang B, Wang JL, et al. (2016) Discussion on the situation of air pollution control and its existing problems and suggestions. Exploration Science, 3: 00301-00301.
[2] Suo JF Liu Y, Zou SB. (2017) Comprehensive semantic similarity algorithm based on
geographic ontology. Journal of Lanzhou University: Natural Science Edition, 53 (1)/February

[3] Gruber T R.(1993) A translation approach to protable ontology specifications [J]. Knowledge Acquisition, 5 (2): 199-220.

[4] Noy N F, Crubezy M,Fergerson R W,et al.(2003) Protege2000:an open-source ontology-development and knowledge-acquisition environment.Annual Symposium Proceedings, 953-953.

[5] Studer R. (2008) Knowledge engineering:Principles and methods.Data &Knowledge Engineering, 25 (1-2): 161-197.

[6] Werner K. (2001) Ontologies in Support of Activities in Geographical Space[J].International Journal of Geographical Information Science, 15 (7): 613-631.

[7] Chen JJ, Zhou CH, Wang JG. (2006) search progress and analysis of geographic ontology. Frontier of Earth Science, 3 (3): 81-90.

[8] Zhou J, Ding QJ, Wu WN, et al.(2013) Research on the Construction of National Thesaurus System in the Network Environment [J]. Library and Information Service, 57 (16), 6-10.

[9] Li B, Liu B, Liu JP, et al. (2015) The research progress review and prospect on Geo-ontology [J]cience of Surveying and Mapping, 40 (4): 53-57.

[10] Lai YX, Li YJ, Liu J. (2013) Construction of Ontology-based Rice Breeding Method Knowledge Base. Journal of Beijing University of Technology, 45 (12) 1181-1197.

[11] Zhou XY, Liu HX, Su H, et al. (2019) Case retrieval method of seed metering device based on ontology and multi-attribute decision making. Journal of Northeast Agricultural University 50 (11): 61-70.

[12] Mu Q, Zhang SQ. (2013) An evaluation of the economic loss due to the heavy haze during January 2013 in China. China Environmental Science, (11): 2087-2094.