Acquiring Stored or Real Time Satellite Data via Natural Language Query

Xu Chen[1,2], Jin Liu[1,*], Xinyan Zhu[3], and Ming Li[3]

[1] State Key Laboratory of Software Engineering, Wuhan University, Wuhan, China
[2] Language Technologies Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, USA
[3] State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan, China
{xuchen,jinliu}@whu.edu.cn

Abstract. With the advent of Sensor Web, the satellite data acquired by sensor systems could be shared among users immediately. Our research has led to an implementation of natural language queries such that users without particular knowledge of satellite imagery can describe easily for what they need. We use a rules-based method to retrieve named entities, with the help of a knowledge base and uses existing Sensor Web services for acquiring stored or real time satellite data. We use rule-based methods to align time, location and domain task entities in natural language queries with Sensor Web services with standard times, geographical coordinates, and satellite attributes. To evaluate our system, we wrote a series of natural language queries in the domains of surveying and mapping, forestry, agriculture, and disaster response. Our queries and satellite data retrieved by the queries were corrected by a group of experts to create a gold standard. Using their remarks as correct, we scored our system results using precision and recall metrics standard for information retrieval. The results of our experiment demonstrate that the proposed method is promising for assisting in Earth observation applications.

Keywords: Observation Task, Satellite Data, Sensor Web, Ontology.

1 Introduction

The core of our research concerns how to bridge the technical parameters of satellite data with natural language queries from users who might be unfamiliar with satellite parameters. More people would like to examine real-time satellite data, or integrate these data into their own services. Archives of satellite data are available through keyword search of geo-portals such as the Google MAP[1], the National Aeronautics and Space Administration (NASA) Earth Observing System Data and the Information System EoDIS[2].

---

* Corresponding author.

[1] https://www.google.com/maps/preview

[2] http://reverb.echo.nasa.gov/reverb/#utf8=%E2%9C%93&spatial_map=satellite&spatial_type=rectangle

L. Chen et al. (Eds.): APWeb 2014, LNCS 8709, pp. 212–223, 2014.

The Sensor Web [2] allows satellite data to be obtained in real time. But how to find it? Sensor Web services integrate easily with other Web services. However, most people find the interfaces of service (such as Sensor Observation Service [12] or Sensor Planning Service [14]) opaque due to their technical parameters and specifications.

In this study, we propose a novel method that allows natural language query to search and retrieve archived or real-time satellite data. We use a rules-based method to find named entities, with the help of a knowledge base. We use rule-based methods to link time, location and domain tasks entered by users with the technical specifications of the existing infrastructure for the Sensor Web.

The remainder of this paper is organized as follows: Section 2 outlines an architecture of our system. A structured natural language template and knowledge base are described in Section 3. Section 4 explains how keywords and named entities are identified. A system implementation and the experimental study are discussed in Section 5. Section 6 describes related work, and Section 7 summarizes the conclusions and gives potential directions for future research.

2 Architecture

Many Sensor Web services have been implemented and can be accessed by Internet, eg., the company 52°North, with its Sensor Web community[3], which can be used as a data layer for acquiring satellite data. We require an intelligent analysis layer as a sort of middleware between client and existing Sensor Web services. Therefore, the framework could be divided into three layers: user interface layer, intelligent analysis layer, and data layer, as shown in Fig. 1.

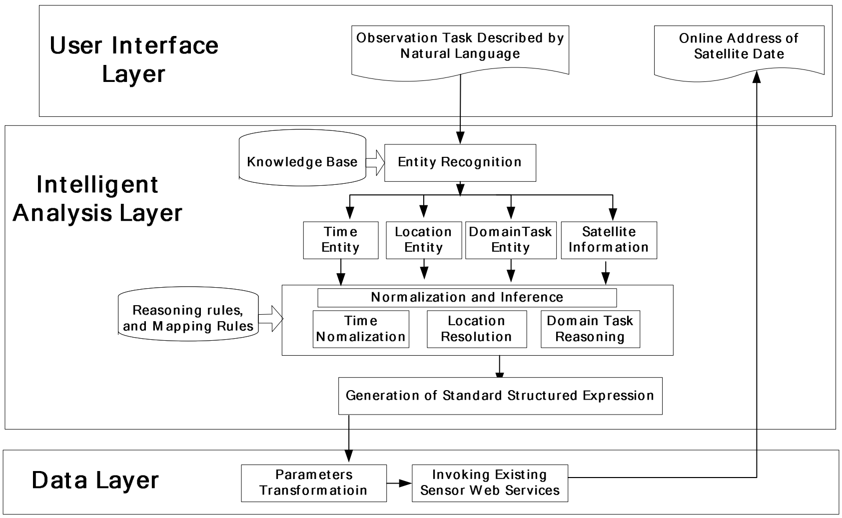

Fig. 1. Acquiring stored or real time satellite data via natural language query

[3] http://52north.org/communities/sensorweb/

User Interface Layer: is a simple and user-friendly Web browser client. Anyone can use it to input some keywords or a natural language sentence for describing an observation task. Multi-condition combination query and a bulk feed are also supported. Results are also shown in this layer.

Intelligent analysis layer: is the core layer that achieves task recognition and reasoning. It includes a rules-based classifier for named entities which draws upon the knowledge base. In section 3.2, we describe how the knowledge base includes a time ontology, location ontology, satellite ontology, domain task ontology. The output of this layer are normalized values based on the format which we defined.

Data Layer: This provides satellite data based on the services layer. A parameter transformation function is provided in this layer which transform the result of Intelligent Analysis Layer to values of interface parameters of the standard Sensor Web services, then we can invoke existing Sensor Web services such as the Sensor Observation Service, or the Sensor Planning Service.

3 A Structured Natural Language Template and Knowledge Base

3.1 Template for Input

In generally, users prefer to describe the observation task by natural language rather than formalized language. However, keyword search of data collections often lack precision, and automated parsing of unrestricted natural language may be inaccurate [16]. In this study, we propose a structured natural language template, to reduce the parsing problems of Natural Language Processing and remove the limitations of keyword search.

We investigate four areas where such satellite data might be useful, including Bureau of Surveying and Mapping, Department of Forestry, Department of Agriculture, Ministry of Civil Affairs. We asked officials who worked in these departments for parameters for what satellite data they might need. We collected more than 200 requirement descriptions were collected. Then we used a ground-up approach to construct a template for future unseen queries based on their requirements. It is based on this experiment that we found that time, location, domain task and satellite requirement are the basic elements of observation task, which we included in our template. These basic elements can be used to construct a data query or a satellite plan. Therefore, we present a structured natural language template, which is defined as follows:

- *ObservationTask* = { *Time*, *Location*, *DomainTask, SatelliteRequirement* }

Time expresses when the task should be executed, eg., About 2003.8.21.

Location expresses where the task covers with, e.g., southwest Montana.

DomainTask describes a specific domain task, e.g., monitoring of wildfires.

SatelliteRequirement expresses the detailed description of image parameters and sensor parameters, e.g., MODIS.

User could easily describe an observation tasks based on this template:

" About 2003.8.21, monitoring of wildfires in southwest Montana, the sensor should be MODIS. "

3.2 Knowledge Base

Ontologies are used to model the domain knowledge, and organize the concepts and properties of time, spatial information, domain task and satellite. In order to recognize the entities in a user query, we collected vocabulary specific to four domains: forestry, agriculture, surveying and mapping, and disaster response. We collect terms, relationships from Wikipedia[4], WenkuBaidu[5], terminological dictionaries and standards of each domain (such as ISO Standards for Geographic Information[8]), website of organizations and institutions (such as ISO/TC211[6]). Then we built a knowledge base with a time ontology, a location ontology, ontologies for our given task domains, a satellite ontology, task reasoning rules and spatiotemporal calculation rules.

Ontologies in the Knowledge Base. Time information describes the time or duration of a user observation task (such as assess the wildfire's area). Location describes the observation location of the task. Domain tasks indicate a specific task in a specific domain, satellite information is about satellite data to achieve observation task. Therefore, we built four types of ontologies to model the requirement of observation task.

Time Ontology: OWL-time is a temporal ontology that provides a vocabulary for expressing data on the topological relations between instants and intervals, together with the duration and date–time information [6].We adopt the OWL-time as a basis for the time ontology. Time ontology includes temporal terms (e.g., festival, season), temporal units (e.g., year, month, week, day), temporal qualifier (e.g., before, after). The Chinese temporal ontology is built in the study additionally, most of concepts refer to Time ontology in English. However, Mandarin has some special temporal concept, the concepts of the Chinese lunar calendar and the traditional solar term are added into the Chinese time ontology, which provide better support for the observation task described in Chinese.

Location Ontology: We reference GeoNames Ontology[7] and build the Location ontology. Location ontology is used to organize concepts of toponym, spatial relationships, feature types, spatial range and so on, which consists of the geo-feature entity, geo-feature-type, and spatial relationship ontologies. The geo-feature entity ontology

[4] https://www.wikipedia.org/
[5] http://wenku.baidu.com/
[6] http://www.isotc211.org/
[7] http://www.geonames.org/ontology/ontology_v3.1.rdf

includes the place name, geocoding, feature type, and footprint. The geo-feature-type ontology is a classification ontology of the feature type. The spatial relationship ontology is built to describe the spatial relationship which defined in the DE-9IM model [3]. Toponyms in gazetteer are regarded as instances of Location Ontology.

Domain Task Ontology: The vocabulary came from speaking with experts and noting keywords about tasks that might use the satellite data, terminological dictionaries and standards of domain. The inter-relations were made by hand with the help of domain experts. In this study, "task" refers to the observation task, especially for Earth observation task. Domain means application domain of Earth observation technology, different domain has different tasks, vocabularies, and concepts, building domain task ontology is a heavy work. The domain task ontology consists of observation object, observation action, object attribute, and their relations.

Satellite Ontology: The vocabulary came from Wikipedia and some satellite websites. The inter-relations were also made by hand with the help of domain experts. Properties of Satellite Ontology include id, mission type, operator, reference system, regime, semi-major axis, eccentricity, period, epoch and so on. Satellite Ontology refer to Sensor Ontology and Satellite Data Ontology. Sensor Ontology describes specific properties of sensor, e.g. orbit, scan rate, swath. Satellite Data Ontology describes specific properties of data, e.g. spatial resolution, fabric width, band, and signal noise ratio.

Reasoning Rules in the Knowledge Base. In addition to these ontologies, we use rules to find enough information in order to determine which satellite to call upon.

Task Reasoning Rules: Many observation tasks described by natural language are incomplete. If the query lacks satellite-related information, for example, it is hard to find satellite data. Therefore, we defined a variety of task requirements based on template in several specific domains (disaster response, agriculture, surveying and mapping, forestry) and encode the reasoning in Semantic Web Rule Language.[8] For example, "monitoring of forest fire" is a typical observation task, which demands satellite data with a spatial resolution of less than 1000 meters and a near-infrared wave band. The rules can be expressed as follows:

Monitoring of Forest fire (?task) → satelliteData(?x) ∧ hasSpatialResolution(?x,?y) ^ swrlb:lessThan(?y,1000) ^hasBand(?x,?z) ^ bandName(?z,?bName)^ swrlb:stringEqualIgnoreCase(?bName,"Near Infrared")

Spatiotemporal Calculation Rules: We also use rules to find a specific time and location. For example, we use spatial analysis calculation, spatial relationship calculation, coordinate transformation (WGS 84 is the first choice), time transformation, normalization of relative temporal expressions (e.g. today) and implicit temporal expressions (e.g. spring season) and so on.

[8] http://www.w3.org/Submission/SWRL/

We have found experimentally that our ontologies, reasoning rules, and gazetteer provide a sufficient resources to support Named Entities Recognition, normalization and inference of user queries in natural language.

4 How Our Recognition Algorithm Works in Details

The observation task described by natural language is processed by a rule-based algorithm that recognizes keywords and named entities with the help of the knowledge base described in section 3.2. The way it works is that time entities, location entities, domain task entities, and satellite requirements are recognized. Then normalization and inference is used to gain deeper understanding of the user query.

Algorithm. A Rule-based Recognition

```
Input : User query Q_i
Output: R{T_i|P_i, L_i|A_i, S_i }
for each Q_i
    Named Entities Recognition(Q_i) ;
    if (time entities exist and NumofTime == 2) then
            TemporalCalculation(T_1,T_2);
                Period of Pi, add Pi to R;
    else if (time entities exist and NumofTime == 1) then
            time normalization T_i;
                add T_i to R;
    if ( location entities exist and 6>NumofLocation>1) then
            SpatialCalculation (L_1,L_2...L_Num);
            MBR of A_i;
             add A_i to R;
    else if (location entities exist and NumofLocation==1) then
            resolution Location L_i;
                add L_i to R;
    if domain task entities exist then
            task reasoning S'_i;
            if satellite requirement exist then
            resolution satellite information S''_i;
        S_i= intersection(S'_i,S''_i) ;
        add S_i to R;
    return R;
```

The above is pseudocode for our algorithm. The input to the algorithm is a user query or set of queries, Q_i. The output is set R, which includes normalization time T_i or a period of time P_i, geographical coordinates of L_i or a minimum enclosing rectangle of observation area A_i, satellite information S_i which is a interaction result of task reasoning S'_i and satellite requirement S''_i.

4.1 Expression Rules

We inferred these rules based on collecting actual data from professionals in our domain fields. We discovered four types of expression rules: for time, location, task and satellite parameters, then we defined these rules based on Backus–Naur Form. Each rule is explained below.

Time. Time data divides into instants and intervals. The terms contain “Day,” “Month,” “Year,” “Christmas,” “August,” and so on. An example of an instant is 2003-08-21 21:00, and an example of an interval is a month. The qualifier for the time contains “before,” “after,” “between,” and so on. These vocabularies are used to describe time in the observation, e.g., “before August 21, 2003,” The expression rules of the temporal information are described as follows:

Time Information::={Qualifier} + <Value> + [Month] + <Value> + [Day] +<Value>+ [Year].

Location. “In the boundary between Montana and Idaho” and “in southwest Montana,” are examples of the location information in the observation task, which include toponym or spatial range, terms of spatial relationship and spatial direction. The expression rule of the spatial information resemble in the following, the Qualifier indicates words that describe the position and direction, e.g., “in,” “at,” “between”. The Relationship indicates the spatial relationship terms, e.g., “across,” “intersect”. Toponym indicates terms such as “administrative division,” “river,” “mountain”, spatial range is another choice to describe a observation range.

Location Information::= [Qualifier] + [Relationship] + {<Toponym> / <Spatial Range> + [Qualifier]}.

Task. Domain task information is the core of the observation task. This is subdivided into task actions (found in natural language in terms such as “monitoring” or “updating”), task aspect (found in natural language queries in terms such as “area,” or “desertification”), and task object (found in natural language in terms such as “wildfires,” or “digital terrain map”).

Domain Task::=<Action> + {[Aspect]} + <Object>.

Satellite Parameters. In the description of the observation task, users impose some restrictions on the satellite parameters (such as “the spatial resolution should be better than 10 m.”). Qualifier indicates the words that express comparison, such as “better than” or “equal.” Value indicates a quantifier to express the parameters, e.g., “5” or “five.” Unit is the unit of the parameters, e.g., “m” or “km.”

Satellite Requirement::=[Satellite parameters] + {Qualifier} + <Value> + <Unit>.

4.2 Normalization and Inference

After we get the named entities of user query, in some cases we still cannot find suitable satellite data because we lack specific time, latitude and longitude, and/or satellite information. Therefore, we execute normalization and inference based on the result of Named Entities Recognition.

For example, the user input this query:

"About 2003.8.21, monitoring of wildfires in southwest Montana, the sensor should be MODIS."

After the Named Entities Recognition, the time entities, location entities, and domain task entities and satellite requirement are extracted based on expression rules and knowledge base. The intermediate result resembles the following:

Time Entity: About 2003.8.21
Location Entity: southwest Montana
Domain Task Entity: monitoring of wildfires
Satellite Requirement: MODIS

Then in another step, which also called normalization and inference. It is used to transform the named entity to standard time information, spatial information, and satellite information, the result as following:

Time information, which is described in a standard time representation:

Start time: 2003-08-21T00:00:00, End time: 2003-08-21T24:00:00.

Spatial information, southwest Montana is transformed to longitude and latitude by spatial calculation:

Latitude: N45°01′57.39″~ N46°40′56.90″, Longitude: W107°12′01.38″~W111°45′38.87″.

Satellite information, which is the intersection of task reasoning results and original satellite requirement, e.g., “monitoring of forest fire” is a typical observation task, which requires sensor with a spatial resolution of less than 1,000 m and a near-infrared wave band, and in this task, the MODIS sensor requirement is expressed definitely, which also satisfies the requirement of monitoring of forest fire. Therefore, the intersection result is:

Satellite information: MODIS.

5 Prototype System and Experiment

5.1 Prototype System

We developed a prototype system to test and verify this method of retrieving satellite data archived or in real time. The Web client of the prototype system is shown in Fig.2. A input field is on the top panel, we can input an observation task described by natural language into the field. After we run the processing function, the time, location, task entities and satellite requirement are extracted by Named Entities Recognition (NER), the result is shown in left panel. Then, we can get result of the normalization and inference, the normalization of temporal information, spatial coordinates, the detailed satellite information are shown on the right panel. Then, we can invoke existing Sensor Web services based on the results to acquire satellite data.

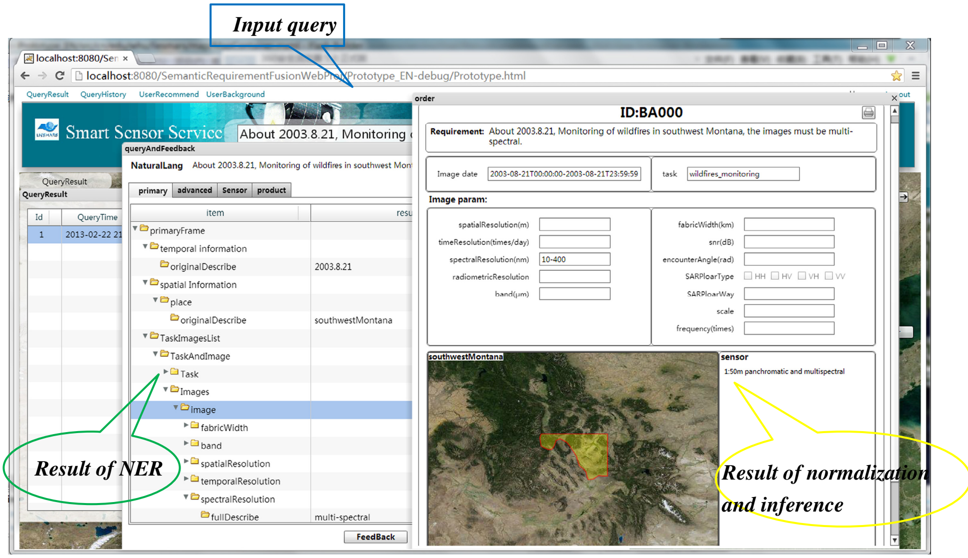

Fig. 2. The web client of prototype system

5.2 Evaluation

We tested the ability of our system to retrieve satellite data based on natural language query. We used observation task queries in the domains of surveying, forestry, agriculture, disaster reduction. We collected 212 observation task keywords in Mandarin described by experts. Then the authors of this paper ourselves wrote 623 more sets of tasks with related, but not duplicate keywords. We had experts re-read the queries that the authors had written to verify their plausibility, and to modify the queries if necessary. The break-down of queries by domain in the sample data set appears in Table1.

We show four examples translated to English as follows:

Sample query_1, from surveying and mapping domain: ***2012, producing Digital Line Graphic of Beijing Changping district, the scale of the product is 1: 460000.***

Sample query_2, from forestry domain: ***In the autumn of 2011, wooded area investigation of Heilongjiang Province, resolution is not less than 10m.***

Sample query_3, from agriculture: ***2011.06.21-2011.07.10, yield monitoring of spring wheat in the north.***

Sample query_4, from disaster response: ***On 2003/8/21, assess the wildfire's area in northeast Idaho, the spatial resolution should be better than 10m.***

To evaluate, we randomly selected six groups queries (each group has 100 queries) from these domain samples. Then we gave both query and results to a group of experts in each domain to make a gold standard. Then we scored the recall and precision, with the results shown in Table 2.

Table 1. Test samples statistics

Domain	Surveying and Mapping	Forestry	Agriculture	Disaster Reduction
#Sample queries	199	210	209	217

We score results in terms of precision and recall:

$$Precision\ of\ NER = \frac{numbers\ of\ correct\ recognized\ query}{numbers\ of\ recognized\ query} \quad (1)$$

$$Recall\ of\ NER = \frac{numbers\ of\ correct\ recognized\ query}{numbers\ of\ query} \quad (2)$$

If one of the named entities in a query is not extracted, we judge the query is not recognized. For example, in Sample query_2, If ***"In the autumn of 2011"*** is not recognized, despite ***"wooded area investigation***, ***Heilongjiang Province***, ***resolution is not less than 10m"*** are recognized correctly, this query is not recognized correct. This represents the first pass of the query through the system. However, this is sometimes insufficient because we lack specific time, latitude and longitude, and/or satellite information. Therefore, we score results of normalization and inference:

$$Precision\ of\ normalization\ and\ inference = \frac{numbers\ of\ correct\ normalization\ and\ inference\ query}{numbers\ of\ normalization\ and\ inference\ query} \quad (3)$$

$$Recall\ of\ normalization\ and\ inference = \frac{numbers\ of\ correct\ normalization\ and\ inference\ query}{numbers\ of\ correct\ recognized\ query} \quad (4)$$

The precision and recall of normalization and inference is based on the results of NER, if one of the named entities is not normalized or reasoning correctly, we judge the query is not understood. For example, if we can't calculate the spatial range of ***Heilongjiang Province*** correctly, this query is not inference correctly, although we get the standard time and satellite specific information.

Table 2. Precision and recall statistics

	1	2	3	4	5	6	Average
Precision of NER	98.3	98.7	96.5	98.9	93.4	96.4	97.03
Recall of NER	96.4	96.3	97.2	95.1	95.6	96.1	96.11
Precision of normalization and inference	96.1	95.6	94.3	96.9	91.1	94.9	94.81
Recall of normalization and inference	95.4	97.8	96.5	97.2	93.7	95.3	95.98

The average precision and recall of NER is above 96%, because we utilize the a large knowledge base and variety expression rules, but a good result relies on the completeness of knowledge base and rules, we provide a function to add new knowledge and rules, to enhance the extendibility of our system. The average precision and recall ration of normalization and inference is also above 94%, assuming that results from NER are correct. Therefore, our prototype system is promising for assisting in Earth observation applications.

6 Related Work

Our system is able to return satellite data to users based on their task domains, while comparable systems by the NASA and the ESRI Company return data only in direct response to user keywords. Recall that the core of our research concerns how to bridge the technical parameters of satellite data with natural language queries from users who might be unfamiliar with satellite parameters. We break this into sub-parameters in order to review the literature.

Location. Recognizing geospatial information in document files (for example, from txt, html, xml and doc) is an active research direction in the geo-spatial domain. Most studies focus on place name, also called toponym recognition. The recognition of toponyms based on Gazetteer [7] is a fundamental method. Many NER approaches are used for toponym recognition, the approach that employs dictionaries and hand-made rules is very popular [9]. The approach based on Machine Learning is another popular method for toponym recognition, including Maximum Entropy[1], Conditional Random Field sequence models [4] and so on. A combination of rule-based method and Machine Learning is a new trend for toponym recognition [10, 5].

Location and Time. Some algorithms combine temporal and geographic information. Strötgen et al. use temporal and geographic information extracted from documents and recorded in temporal and geographic document profiles[15].

Location, Time and Task. Events happen at a given place and time[11]. Most of the research focuses on identifying events from temporally-ordered streams of documents and organizes these documents according to the events they describe[17]. Our observation task is a special type of event, which is more closely linked to temporal and spatial information than other events.

Research indicates that even state-of-the-art entity recognition systems are brittle, meaning that they are developed for one domain but do not typically perform well on other domains [13]. Therefore, research on entity recognition to acquire satellite data is significant.

7 Conclusion

In this study, we describe a novel method to acquire stored or real time satellite data via a natural language query based on existing SWE services. Based on users input, our algorithm uses Named Entities Recognition, normalization and inference to find relevant items in the satellite data. Our evaluation is based on our in-house prototype system which showed precision and recall. Open questions for future research include how to expand the reasoning rules that match between users input and ontologies of data automatically by machine learning.

Acknowledgment. This work was supported by the grants of the National Basic Research Program of China (2011CB707101), the National Natural Science Foundation of China (41201405, 61070013), China Scholarship Council No. 201308420300.

References

1. Bender, O., Och, F., Ney, H.: Maximum Entropy Models for Named Entity Recognition. In: Proceedings of the 7th Conference on Natural Language Learning (CoNLL 2003), Edmonton, Canada, pp. 148–151 (2003)
2. Delin, K.A., Jackson, S.P.: The Sensor Web: A New Instrument Concept. In: Proceedings of the SPIE's Symposium on Integrated Optics, San Jose, CA, vol. 4284, pp. 1–9 (2001)
3. Egenhofer, M.J., Franzosa, R.D.: Point-set topological spatial relations. International Journal of Geographical Information Systems 5(2), 161–176 (1991)
4. Finkel, J.R., Grenager, T., Manning, C.: Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, Ann Arbor, MI, pp. 363–370 (2005)
5. Gelernter, J., Zhang, W.: Cross-lingual geo-parsing for non-structured data. In: 7th Workshop on Geographic Information Retrieval (GIR) Orlando, Florida, USA, pp. 64–71 (2013)
6. Hobbs, J.R., Pan, F.: An Ontology of Time for the Semantic Web. ACM Transactions on Asian Language Processing: Special issue on Temporal Information Processing 3(1), 66–85 (2004)
7. Hill, L.: Georeferencing – The Geographic Associations of Information, pp. 92–154. MIT Press, Cambridge (2006)
8. Kresse, W., Fadaie, K.I.: Standards for Geographic Information. Springer (2004)
9. Leidner, J.L., Lieberman, M.D.: Detecting Geographical References in the Form of Place Names and Associated Spatial Natural Language. The SIGSPATIAL Special 3(2), 5–11 (2011)
10. Lieberman, M.D., Same, H.: Multifaceted Toponym Recognition for Streaming News. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, Beijing, China, pp. 843–852 (2011)
11. Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J.: Introduction to wordnet: an on-line lexical database. International Journal of Lexicography 3, 235–244 (1990)
12. Na, A., Priest, M.: OGC Implementation Specification 06-009r6: OpenGIS Sensor Observation Service. OpenGIS Sensor Observation Service. Open Geospatial Consortium Inc., Wayland (2007)
13. Poibeau, T., Kosseim, L.: Proper Name Extraction from Non-Journalistic Texts. In: Computational Linguistics in the Netherlands Meeting, Amsterdam, New York, pp. 144–157 (2001)
14. Simonis, I., Echterhoff, J.: OGC Implementation Specification 09-000: OpenGIS Sensor Planning Service. Open Geospatial Consortium Inc., Wayland (2011)
15. Strötgen, J., Gertz, M., Popov, P.: Extraction and Exploration of Spatio-temporal Information in Documents. In: Proceedings of the 6th Workshop on Geographic Information Retrieval (GIR 2010), Zurich, Switzerland, pp. 1–8 (2010)
16. Tamand, A.M., Leung, C.H.C.: Structured Natural-Language Descriptions for Semantic Content Retrieval of Visual Materials. Journal of the American Society for Information Science and Technology 52(11), 930–937 (2007)
17. Vavliakis, K.N., Symeonidis, A.L., Mitkas, P.A.: Event identification in web social media through named entity recognition and topic modeling. Data and Knowledge Engineering 88, 1–24 (2013)