Generation of Multilingual Personalized Environmental Bulletins from an OWL-based Ontology

Nadjet Bouayad-Agha¹, Gerard Casamayor¹, Simon Mille¹, Marco Rospocher², Luciano Serafini², Jürgen Moßgraber³, and Leo Wanner¹,⁴

Abstract

In this paper, we tackle the problem of generation of user-oriented multilingual environmental information from ontologies in the context of a personalized environmental decision support service. We present a unified multiple layer ontology framework modeled in OWL that consists of three ontology layers: the domain ontology, the domain communication ontology, and the communication ontology. The domain ontology contains factual application-neutral concept configurations and relations. The domain communication ontology models data aggregation, qualitative interpretation of numerical data, user tailored warnings and recommendations triggered by an environmental condition given in a specific context, etc., while the communication ontology specifies knowledge needed for the tasks involved in the generation process, and is populated using a pipeline of SPARQL queries. We show how a large scale instantiation of this framework in the environmental domain serves multilingual NLG.

1. Introduction

In this paper, we tackle the problem of generation of user-oriented multilingual environmental information from ontologies in the context of a personalized environmental decision support service.

Text generation from ontologies (or more generally, knowledge bases, KBs) is a common research topic in Natural Language Text Generation (NLTG). However, until recently, the proposals tended to have in common that they start from KBs of limited size which either already contain linguistically-oriented knowledge structures as, e.g., the Upper Model [1] or the MIAKT ontology [2], or explicitly assign to the conceptual knowledge structures their possible linguistic realizations; see, among others, [3–6]. In other words, they either intermingle application-neutral concept configurations with application-oriented configurations, which is prohibitive when the same KB serves as resource for several different applications, or they require manual intervention to adjust the linguistic realization when the KB is extended by new configurations, which is prohibitive if the generator is supposed to scale up.

To address this problem, [7] suggest to extend an application-neutral base ontology (in their case, soccer ontology) by a second layer (they call extended ontology) in which the required linguistically-oriented configurations are introduced by means of off-line inferences. Their proposal facilitates a clean separation of conceptual and linguistic structures, but the off-line creation of the extended ontology is a serious drawback if the base ontology is dynamic, i.e., instantiated with new concept instances, for instance, in the course of time or depending on the query of a user. Furthermore, they define the text planning knowledge structures outside the ontology model itself, which makes the mapping of the ontology structures onto text plans more complex and cumbersome than necessary and negatively affects the maintenance and portability of the generator.
In what follows, we present a multiple layer ontology framework which further improves on [7]’s proposal and illustrate its application to the generation of multilingual user-oriented environmental reports. The framework consists of three ontology layers that reflect [8]’s distinction of the three types of knowledge needed by NLTG: the domain ontology, the domain communication ontology, and the communication ontology. The domain (or base) ontology contains factual application-neutral concept configurations and relations. The domain communication (or extended) ontology models data aggregation, qualitative interpretation of numerical data, user tailored warnings and recommendations triggered by an environmental condition given in a specific context, etc., while the communication ontology, which specifies knowledge needed for the tasks involved during content selection and discourse structuring, and is populated using a pipeline of SPARQL queries, contains text planning relevant knowledge structures.

2. Multi-layered ontology as a starting point for generation
The three layers of our knowledge representation model are implemented in OWL [9], the state of the art Semantic Web ontology language. In what follows, we sketch the main features of each of them.

2.1 The domain ontology
The domain ontology models domain-specific knowledge relevant to the considered application settings of a personalized environmental decision support service that collects its data from the web. It captures thus the concepts, relations, and individuals related in particular to (i) environmental data: datasets (e.g., temperature, wind speed, birch pollen count, CO2 concentration, etc.), which can concern observed, forecasted, historical, and/or statistical data and be qualitative or quantitative; (ii) environmental measurement providers: properties of the web-based providers of environmental data measurements (e.g., meteorological web-sites, air quality monitoring stations); (iii) geographical information: the area for which a certain environmental data measurement holds or the area relevant to the user request; (iv) environment-related user requests: the request selected by the user among those supported by the system; (v) environment-affected user profile aspects: user age, user sensitiveness to some environmental conditions (e.g., birch pollen), user diseases related to environmental conditions (e.g., asthma).

Currently, the domain ontology comprises 216 classes, 136 datatype and object properties, and 582 individuals. For the construction of the ontology, techniques for automatic ontology extension are used [10]. Each time a decision support request is submitted to the system, the base ontology is instantiated, in successive steps, with content adequate to fulfil the user request. First, a complete description of the user request is instantiated in the base ontology. This description includes the type of request, the profile of the user involved in the request, and the time period and geographical location for which the request applies. From the request, the system determines using Description Logics (DL) reasoning the raw environmental data that are to be retrieved from web-based environmental data providers and distilled into the domain ontology.

2.2 The domain communication ontology
The domain communication ontology contains additional personalized content, spanning from data aggregation (e.g. minimum, maximum, and mean value aggregation of data, computed over the time-period considered in the request), qualitative scaling of numerical data, user tailored recommendations and warn-

---

5 We just briefly sketch the main steps of the instantiation of the base ontology, omitting all the details. For a more comprehensive description of the system workflow, see [11]
ings triggered by the environmental data inserted in the ontology, as well as logico-semantic relations between environmental information (e.g., an implication relation between pollen rating and a recommendation in case of abundant pollen levels). The computation of this inferred content is performed by the decision support module by combining some complementary reasoning strategies, including DL reasoning and rule-based reasoning. A two-layer reasoning infrastructure is currently in place. The first layer exploits the HermiT reasoner [12] for the OWL DL reasoning services. The second layer is stacked on top of the previous layer. It uses the Jena [13] RETE rule engine, which performs the rule-based reasoning computation.

The domain communication ontology comprises 77 classes, 34 datatype and object properties, and 249 individuals. 221 rules have been defined for inferring new individuals from the base KB.

2.3 The communication ontology

The communication ontology models the concepts and relations needed for the two text planning modules in generation: content selection and discourse structuring.

As is common in report generation, our content selection is schema (or template)-based. Therefore, the communication ontology defines a class Schema with an n-ary schema component object property whose range can be any individuals of the domain and domain communication KBs. Similar to [7], we assume the output of the discourse structuring module to be a well-formed text plan which consists of (i) elementary discourse units (EDUs) that group together individuals of the domain and domain communication ontologies, (ii) discourse relations between EDUs and/or individuals of the domain and domain communication ontologies, (iii) sentence units that group together EDUs to be realized in the same sentence, and (iv) precedence relations between sentence units. This structure translates in the communication ontology into three top classes: Sentence with an n-ary sentence component property and a linear precedence property, EDU with an n-ary EDU component relation, and Discourse Relation with nucleus and satellite relation.

![Sample Text Plan](image-url)
Figure 2 shows an instantiation of the discourse structuring concepts and relations in the environmental domain for pollen rating and associated warning and recommendation messages. In essence, this is an output text plan. A set of SPARQL query rules are defined to instantiate the various concepts and relations. Currently, the communication ontology comprises 13 classes and 40 object and data properties.

3. Ontology-based natural language report generation

The architecture of our report generator is a standard pipeline architecture ‘text planning \(\rightarrow\) linguistic generation’ [15], with text planning being internally subdivided into content selection and discourse structuring.

3.1 Content Selection

Content Selection (CS) operates on the output of the decision support module, which populates the domain and domain communication ontologies with the knowledge relevant to the submitted user query and profile. It selects and groups by topic the content to be included in the report. The following topics are treated:

- Location and time chosen by user.
- Air Quality (AQ) related information (AQ index minimum and maximum values and ratings, responsible pollutant(s), warnings and recommendations for the AQ, and any LSRs between the schema’s components).
- Individual pollutant related information (one schema per pollutant with value, rating, warnings and recommendations, and any LSRs between the schema’s components).
- Pollen related information (pollen minimum and maximum counts, warnings and recommendations related to these counts, and any LSRs between the schema’s components).
- Weather conditions (one schema per weather phenomenon (rain, temperature, windspeed, humidity, sky condition) that includes any of: weather minimum and maximum values and ratings, warnings and recommendations for these weather values and associated LSRs).
- Threshold exceedance of individual pollutant values, with associated warnings, recommendations and corresponding LSRs.
- Missing data (one schema per missing data).

For selection, topic-specific queries are used. The inclusion of a given individual in a schema can be subject to some restrictions defined in the queries; for example, if the minimum and maximum AQ index ratings are identical, then only the maximum value is selected.

3.2 Discourse structuring

Discourse structuring (DS) is carried out by a pipeline of four rule-based submodules: (i) Elementary Discourse Unit Determination (EDU-D), (ii) Mapping LSRs to discourse relations (M-LSR-DR), (iii) Sentence Unit Determination (SUD), and (iv) Sentence Unit Ordering (SUO).

EDU-D groups topically related individuals into propositional units starting from the schemas determined during CS; Figure 2 shows two EDUs identified in the pollen-related schema, one for the pollen count and another for the recommendation message. M-LSR-DR maps LSRs which hold between individuals of the domain communication ontology for discourse relations [14] to be realized intra-sententially) or be-

---

6 For sake of simplification, EDU and Sentence components are represented using a “bubble” rather than a component object property from each EDU/Sentence to each of its components.
tween their EDUs (for relations to be realized intersententially). In Figure 2, the name of the discourse relation of the individual \textit{rel\_map\_a1}, namely Evidence is specified as property and therefore not visible. SUO introduces a precedence relation between sentence units using the following heuristics determined by domain communication experts:

1. Always start with the location and time of the user query.
2. Present exceedance-related warnings and values first.
3. Next present any non-exceedance-related warnings, recommendations and values.
4. Information on any missing data is placed last.
5. For a given topic, present the minimum value/rating before the maximum.
6. Use the following partial order to order sentence units at the same depth of the structure: AQ > pollen > weather.
7. For weather information, use the following (partial) order: temperature > wind speed > rain > sky condition > humidity

3.3 Linguistic generation

Our linguistic generation module is based on a multilevel linguistic model of the Meaning-Text Theory (MTM) [16], such that the generation consists of a series of mappings between structures of adjacent strata (from the conceptual stratum to the linguistic surface stratum). Starting from the conceptual stratum, for each pair of adjacent strata, a transition grammar is defined. For details, see [17].

The generator receives as input a Conceptual Structure (ConStr) in the sense of [18], derived from the text plan produced by the text planning module. In a sense, ConStr can thus be considered a projection of selected fragments of the ontologies onto a linguistically motivated structure. ConStrs are language-independent and thus ideal as starting point of multilingual generation. Figure 3 shows a sample ConStr derived from the text plan in Figure 2. The main difference with the text plans is that the ConStr must contain all the information mentioned in the sentence, and only this information. As a result, some concepts, such as \textit{measurement}, implicit in Figure 2, must be made explicit since the final sentence needs to mention that there was some kind of measurement (in the case of birch pollen, it is \textit{count}). On the other hand, some nodes from the text plan which only represent domain-specific knowledge are simply not mapped to the conceptual structure. Grammatical and lexical resources used for all successive mappings are detailed in [17].

Figure 4 shows, for illustration, a sample bulletin in English, Finnish, and Swedish generated from the instantiated ontology for one day.
4. Evaluation

Given that the domain communication ontology provides the bulk of the content from which the generation starts, we evaluated both the domain communication ontology population techniques and the generation techniques.

Firstly, we performed a qualitative evaluation of the techniques for reasoning and user tailored decision support that populate the domain communication ontology. The evaluators were asked to judge the appropriateness and completeness of the content instantiated for four different user requests. A 5-value scale was proposed for the evaluation of both appropriateness and completeness. For each evaluator, we determined the percentage of appropriate content produced according to the expert, as well as its completeness (according to the percentage values associated to the values of the completeness scale). The results obtained show an average appropriateness of 90% (with a standard deviation of 25%) and an average completeness of 87% (with a standard deviation of 23%).

Secondly, for the evaluation of content selection, six queries were identified which yield representative contents in the ontology prior to the execution of CS. The content of the ontology after the execution of the DS and right before the execution of the CSS were dumped into a file for each query. For each of the dumps, all the possible instances of the content templates used by the CSS were identified and listed in a readable format. Two environmental expert users were then asked to choose what instances of the templates they considered relevant given the problem description, which includes the user profile, the request, geographic area of the query and dates. Their answers constitute a gold standard of content plans against which other plans can be evaluated. For evaluation purposes, a “majority class” baseline strategy for content selection was devised, whereby instances of each template are selected if more than half of the instances of that same template are also selected by the experts in the gold standard. Table 1 summarizes these measures, repeated for the baseline and the CSS.
Finally, following [20] we performed a two-panel evaluation of the text planning. We selected six content plans with enough variety generated by the Content Selection module. We then had two domain experts produce texts from these content plans in English so as to have two texts per content plan; and one expert to produce texts in Finnish. The twelve texts in English and six texts in Finnish thus obtained constitute the “gold standard set”. We also generated texts in English and Finnish from these same six content plans by the IPS, thus obtaining the “generated set”. Finally we drafted English and Finnish texts from a fixed canned text template, thus obtaining the “baseline set”. The results of the evaluation for Finnish and English by 6 evaluators are depicted in Table 2. For Finnish, but not for English, on average across all the criteria, the generated texts are rated markedly better than the baseline. This is mainly due to the intricacies of the Finnish language (see the intelligibility criterion), which are hard to get with template-based techniques. Our evaluation results in two very different languages allow us to contrast the benefits and limitations of template-based generation and show the advantages of a more linguistic approach.

5. Conclusions and future work

We presented briefly a three-layer ontology framework and have shown how this framework benefits NLTG in general and the production of environmental information in natural language in particular. The ontologies, their user query-based population and the mentioned NLTG modules are implemented in a service-based architecture described in [19,11]. The proposed multilayered OWL ontology-driven NLTG has the advantages that it supports dynamic population of application neutral ontologies, clean separation of domain, domain communication and communication knowledge and codification of text planning tasks in SPARQL queries to populate the communication ontology. Some of our future work includes facilitating the use of different sets of text planning rules (and hence realizations) according to different user profiles, deriving content selection rules from environmental experts interacting with a learning agent, and performing a bigger evaluation with more (varied) texts (for both English and Finnish).

Bibliography


---

Table 1
Content Selection Evaluation with two annotators (A1 and A2)

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td>A1: 0.94</td>
<td>A1: 0.93</td>
</tr>
<tr>
<td></td>
<td>A2: 0.94</td>
<td>A2: 0.94</td>
</tr>
<tr>
<td></td>
<td>Both: 0.94</td>
<td>Both: 0.94</td>
</tr>
<tr>
<td>Precision</td>
<td>0.78</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.76</td>
</tr>
<tr>
<td>F-score</td>
<td>0.85</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>Inverse recall</td>
<td>0.64</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>0.56</td>
<td>0.76</td>
</tr>
<tr>
<td>Inverse precision</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>0.89</td>
<td>0.86</td>
</tr>
<tr>
<td>Inverse F-score</td>
<td>0.75</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>0.69</td>
<td>0.81</td>
</tr>
<tr>
<td>Average F-score</td>
<td>0.80</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Table 2
Average rating by criterion and text set for Finnish (top half) and English (bottom half)

<table>
<thead>
<tr>
<th>Texts</th>
<th>Intelligibility</th>
<th>Packaging</th>
<th>Ordering</th>
<th>Accuracy</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold</td>
<td>4.7</td>
<td>4.6</td>
<td>4.8</td>
<td>4</td>
<td>4.5</td>
</tr>
<tr>
<td>Generated</td>
<td>3.9</td>
<td>4</td>
<td>4.3</td>
<td>4.4</td>
<td>4.2</td>
</tr>
<tr>
<td>Baseline</td>
<td>3.6</td>
<td>3.2</td>
<td>3.8</td>
<td>4.3</td>
<td>3.7</td>
</tr>
<tr>
<td>Gold</td>
<td>4.6</td>
<td>4.3</td>
<td>4.3</td>
<td>4.3</td>
<td>4.4</td>
</tr>
<tr>
<td>Generated</td>
<td>4.3</td>
<td>3.9</td>
<td>4.3</td>
<td>4.5</td>
<td>4.2</td>
</tr>
<tr>
<td>Baseline</td>
<td>4.3</td>
<td>3.8</td>
<td>4.3</td>
<td>4.5</td>
<td>4.2</td>
</tr>
</tbody>
</table>