

The State of Ontology Pattern Research

Karl Hammar

Jönköping University
P.O. Box 1026
551 11 Jönköping, Sweden
haka@jth.hj.se

Abstract. Semantic web ontologies have several advantages over other knowledge representation formats that make them appropriate for information logistics architectures and applications. However, the construction of ontologies is still time-consuming and error prone for practitioners. One recent development that aims to remedy this situation is the introduction of ontology design patterns, codifying best practices and promoting reuse. This paper presents a literature survey into the state of research on ontology patterns, and suggests the use of such patterns for modeling information demand and distribution.

Key words: Ontologies, Ontology design patterns, Literature review

1 Introduction

Semantic web ontologies show promise as a knowledge representation format for use in various applications of information logistics and knowledge supply. Such ontologies allow organizations to define formally how they view their information, in turn enabling harmonization of information systems across the organization. Systems can be built by engineers using ontologies as specifications, or in some cases, ontologies can be directly applied as concrete artifacts in systems, defining schemas or formats of information.

Ontologies have several technical advantages over other types of data models or knowledge representation languages - they are flexible and easily accommodate heterogeneous data, they are platform and programming-language independent, and being based on description logics they can easily be computed on by classifier software, allowing for the inferencing of new knowledge based on that which is already known. This computability capability can also help ensure the consistency and quality of information encoded in ontology languages. Uses of ontologies in information logistics range from competence modeling [1] to requirements management [2] to general knowledge fusion architectures [3].

Ontology patterns were introduced by Blomqvist and Sandkuhl in 2005 [4]. Later the same year, Gangemi presented his work on *ontology design patterns* [5]. Such patterns, encodings of best practices, were intended to reduce the need of extensive experience when developing ontologies. Since then, the latter work's perspective on ontology design patterns has become widespread, and a commu-

nity has formed¹ based on the further developments of these ideas presented in the NeOn project [6].

While patterns have been used in computer science for at least fifteen years (see for instance software design patterns [7], analysis patterns [8], data model patterns [9], software architecture patterns [10, 11], etc.), the use of patterns in ontology engineering is thus a quite new development. As this field has grown, it has now developed enough that an overview and review of the state of research is warranted. In a previous study [12], the author performed such a review of published literature. This work indicated that much work remains to be done on pattern development, evaluation and identification methods, and that the validation applied to ontology pattern research is lacking. In this expanded study, the same method is applied to a larger set of source papers in order to attempt to confirm the findings, and consequences of this situation for the use of ontology patterns in information logistics are discussed.

The remaining part of the paper is structured as follows: after an introduction to the research questions and description of study design (Sections 2 and 3 respectively), the paper presents some key results of the survey (Section 4). Section 5 analyses the results, attempting to answer the research questions posed, and Section 6 concludes and indicates possibilities for future work.

2 Research Questions

To decide on research questions, the author considered the utility of a literature survey. Such a survey's primary purposes include informing the community of what research is being done (and consequently, what is not being and should be done), how it is being done (possibly indicating room for improvements in method), and how the results can be applied. Accordingly, the following research questions are stated:

1. What kind of research on ontology patterns is being performed?
2. How is research in ontology patterns being done?
3. How can existing research into ODPs be used to support information logistics?

3 Method

The method used to perform the review consisted of first finding a set of relevant research papers (that were not already covered by the previous study), selecting the papers from that set mentioning ontology patterns, and then classifying them and extracting from them the data needed to answer the research questions. Since this survey adds a new measure not present in the previous work, the old dataset then had to be updated before the merged results could be analyzed. The following sections discuss the steps in more detail.

¹ <http://www.ontologydesignpatterns.org>

3.1 Selection

The new documents studied in this survey originate from the EKAW and K-CAP conference series and associated workshops, as well as the Journal of Applied Ontology and Journal of Web Semantics. Timespan-wise the delimitation has been to stay within the 21st century, on the intuition that ontology patterns are not likely to be common in literature prior to the advent of the Semantic Web in 2001. Extending the studied period in this manner (the old dataset covered only 2005-2009) also allows inclusion of some ESWC and ISWC conferences not included in the previous study.

In order to find the subset of papers dealing with ontology patterns, all of these papers were subjected to a full-text search. All papers containing the phrases *ontology patterns* or *ontology **word** patterns* (***word*** denoting any one single word) were marked for further manual analysis.

3.2 Initial classification

ODP Importance Classification In order to learn how important the use of or research into ontology patterns is to each particular paper, the section of the paper in which patterns were mentioned was recorded. The intuition was that this information would give an indication as to whether ontology patterns were considered an essential core part of the research (warranting inclusion in the title or abstract) or not.

Content Classification In order to study what type of ODP research is being performed the papers were read and categorized by the author based on how they contributed to ODP pattern research. In order to remain compliant with the review in [12], the same categories and definitions were used. As a small validation of the categories and the tagging procedure, a random set of ten papers were distributed among two researcher colleagues of the author with experience in ontology design patterns (five each). Each colleague then tagged the papers using the same categories and definitions used by the author. 80 % of the papers were tagged identically.

3.3 Extraction

Three types of information was extracted from the document set: metadata about the papers, what usages for ontology patterns were described, and a classification of how well the presented work was validated and tested.

Usage classification While the content classification performed in section 3.2 indicates how the studied papers contribute to ontology pattern research, it does not fully capture the uses to which patterns are put. In order to capture this aspect of research, each document was assigned to a category based on what type of pattern use was featured most prominently in the paper.

Validation Classification In order to survey how ontology pattern research is validated, two procedures were followed. To begin with, the papers were categorized according to in what manner validation or testing of the proposed ideas and theories had been performed. For this purpose, the following four categories were used:

- No validation - there is no mention of any validation of the ideas presented.
- Anecdotal validation - the paper mentions that the research has been validated by use in an experiment or in a project but it provides no further detail.
- Theoretical validation - a validation of features or qualities is performed without any empirical data (including feature comparisons and validations by example)
- Empirical validation - some sort of experimental procedure, case study, or other empirically grounded validation has been performed.

Each paper was assigned to one validation category only. In the cases where a paper matched more than one category, the category mapping to a higher level of validation was selected, i.e. empiricism trumps theory which in turn trumps anecdote.

Having categorized the papers by validation technique, a further study of validation quality was performed against the papers categorized as belonging to the *Empirical validation* group. For this study the metrics and corresponding measurement criteria pertaining to validation developed in [13] were used.

4 Results

The processes described in section 3 provided a large amount of data to analyze, a subset of which is presented in this chapter. The full dataset is too large to include in full, but is available for download ².

Table 1 contains the results of the process described in section 3.2, that is, the labels denoting categories of pattern-related research and the number of papers tagged with each such label. Table 2 adds the usage measure as mentioned in section 3.3. Note that this last table does not include the two papers that could not be assigned to any of the defined usage categories.

Tables 3 and 4, finally, present the results of the validation classification performed in section 3.3, i.e. how the results were validated and, in the case of empirical procedures being used for this purpose, how well the experiments or case studies were described.

5 Analysis and Discussion

In this section the data resulting from the survey is analyzed, attempting to answer the research questions posed in section 2.

² <http://hem.hj.se/~haka/data/ilog2011/ilog2011-dataset.xls>

Table 1. Classification of the reviewed papers' contribution to ODP research.

| Classification | Conferences | Workshops | Journal | Sum |
|--------------------------|-------------|-----------|---------|-----|
| New pattern presented | 13 | 22 | 2 | 37 |
| Patterns used | 14 | 15 | 4 | 33 |
| Pattern usage method | 12 | 14 | 0 | 26 |
| Pattern features | 3 | 7 | 0 | 10 |
| Pattern languages | 5 | 4 | 0 | 9 |
| Pattern typology | 4 | 3 | 0 | 7 |
| Evaluation | 3 | 2 | 0 | 5 |
| Pattern creation methods | 3 | 2 | 0 | 5 |
| Antipatterns | 1 | 2 | 0 | 3 |
| Pattern identification | 0 | 2 | 0 | 2 |

Table 2. Classification of ODP usages.

| Usecase | Conferences | Workshops | Journal | Sum |
|-------------------------|-------------|-----------|---------|-----|
| Feature identification | 2 | 4 | 0 | 6 |
| Ontology Engineering | 23 | 30 | 3 | 56 |
| Ontology Learning | 4 | 4 | 1 | 9 |
| Ontology Matching | 3 | 5 | 0 | 8 |
| Ontology Transformation | 4 | 4 | 0 | 8 |

Table 3. Validation levels of reviewed papers.

| Validation | Conferences | Workshops | Journals | Sum |
|------------|-------------|-----------|----------|-----|
| None | 8 | 11 | 1 | 20 |
| Anecdote | 4 | 4 | 0 | 8 |
| Theory | 12 | 24 | 1 | 37 |
| Empiricism | 13 | 9 | 2 | 24 |

Table 4. Quality of empirical validations.

| Quality indicator | Weak | Medium | Strong |
|--------------------------|------|--------|--------|
| <i>Conference papers</i> | | | |
| Context description | 7 | 5 | 1 |
| Study design description | 1 | 6 | 6 |
| Validity description | 8 | 5 | 0 |
| <i>Workshop papers</i> | | | |
| Context description | 6 | 3 | 0 |
| Study design description | 1 | 3 | 5 |
| Validity description | 7 | 2 | 0 |
| <i>Journal articles</i> | | | |
| Context description | 1 | 1 | 0 |
| Study design description | 0 | 1 | 1 |
| Validity description | 1 | 1 | 0 |

5.1 What kind of research on ontology patterns is being performed?

[12] showed that the areas of pattern creation, identification, evaluation, and antipatterns were least common in the studied papers. This observation holds true also after expanding the dataset as these tags remain the least frequently used. [12] presented two possible interpretations of this fact: either it is a sign of a mature field in which these certain areas have already been explored to their full potential, or it is the opposite, a sign of a young field that has yet to begin formalizing infrastructure for evaluation and best practices for pattern generation or identification. In view of the expanded dataset covering a longer time period and still not covering these areas, it appears to be more likely that the latter is true.

This is unfortunate. Patterns could prove very useful in solving information logistic problems. The reuse of known best practices would likely lead to both a higher quality of the resulting information models/formats and easier interoperability of different information sources by way of shared conceptualizations across domains and problems. However, for this to occur such best practices need to be discovered and patterns formalized based upon them, which appears to be uncommon yet.

Studying the new metric of what patterns are being used for, it is clear that Ontology Engineering (i.e. manual ontology creation and development) is by far the most common use of ontology patterns. It is however interesting to note that 20 of the 31 papers in which patterns are used for other purposes have been published in 2009-2010, possibly indicating that ontology pattern usage is diversifying. If this is the case, it further reinforces the need for developing the type of work mentioned above (evaluation, creation methods, etc.) that is as of yet underdeveloped.

5.2 How is research in ontology patterns being done?

As was touched upon in section 5.1, the type of research on patterns that is taking place seems to be diversifying in the last few years, with new usages being covered to a greater degree. Looking at the content classification one can also see some changes over time. The number of categories that the average 2008-2010 paper is tagged with (1.75) is higher than the average for the entire period (1.58), whereas the number is quite a lot lower for the average pre-2008 paper (1.16). This indicates that the papers in the latter part of the studied period seem to be broader in their use of and research on ontology patterns than the earlier papers. This is perhaps a normal phenomenon in a developing field.

In order to gauge how scholars are cooperating in their work on ontology patterns, one can look at the number of authors and institutions typically involved in the production of a research paper. Of the subset of 80 papers that were written by more than one author, 42.5 % list more than one affiliated institution and 16.3 % lists three or more institutions.. These numbers seem to indicate a quite healthy degree of cooperation between research institutions in the field.

The results indicate that ODP work is primarily taking place at European institutions. All of the top seven institutions counted by number of publications are located in mainland Europe and the UK. Out of a total of 80 institutions that had published, only 13 were located outside of Europe. Out of these 13, 10 were based in the USA and one each based in Australia, Canada, and New Zealand.

With regards to the validation and testing of ODP research, the findings reinforce the point made in [12] that there may be some work to be done. Nearly one third (31.5 %) of the papers published at full conferences contain no validation or only anecdotal validation of the presented work. Another 41.5 % validates the work by way of theory or examples, but provide no real-world testing to ensure validity. Of the papers that do contain empirical testing, it is uncommon to see discussions on the limits of validity of said testing. This situation may be somewhat problematic. Though not all types of research invite the opportunity to perform experiments or case studies, nor actually require them, quite a few papers in the dataset could have benefitted from a more thorough testing procedure.

Finally, looking at how central ontology patterns are to the content of the papers, it was found that the most common situation is actually that patterns are mentioned already in the title, indicating that they are quite central to the papers. One possibility is that this indicates that ontology patterns are primarily used within a small research community and thus written about by people who consider them to be important enough to warrant inclusion in the title.

5.3 How can ODP research be harnessed to support information logistics?

A core issue in achieving working information supply is the capture and modeling of information demand, i.e. the formalization of what information individual coworkers or business units need to have in order to fulfill their business roles. Several methods exist that aim to model information demand (user profiles, situation-based models, context-based models, etc) [14]. All of these methods build upon the idea that there is some type of standard information demand model or theory that can be reused (possibly heavily adapted) in an information supply situation. Ontologies could be used to model such information demand, and their inferencing capabilities used to aid in classifying and distributing incoming information. Frequently occurring information demand models could then be isolated and packaged as reusable ODPs allowing easier transfer of best practice between systems and organizations, not only in terms of information schemata but also in terms of information demand and distribution. These information demand ODPs would in a sense be meta-patterns - not formalizing the information itself, but rather the information about whom the information concerns, and who may have use of it.

Neither the studied papers nor the ODP portal ³ contain patterns such as the ones discussed here. However, there are some published patterns that would likely be useful as building blocks in this type of modeling, for instance "Information realization"⁴, "Time indexed participation"⁵, "Agent role"⁶, etc. Existing work on notation and methods for information demand analysis [15] could also be used as a starting point for such modeling.

Finding and encoding such patterns into ODPs may be more difficult than isolating information content-oriented ODPs since it would require deeper analysis of the organizations where the pattern is found. However, it is a type of pattern development that may be very fruitful and worthwhile, as it is likely that any such patterns found would be less domain-bound than other knowledge/ontology patterns and therefore applicable to a larger number of organizations.

6 Conclusions and Future Work

In summary: ontology pattern scholars are cooperating well and the volume of material published on ontology patterns is increasing yearly. Cooperation is more common on workshop papers than on full conference papers. The community is primarily based in Europe, though some work is also done in US-based organizations. In the last few years pattern usage seem to have diversified. Although the way in which patterns are researched is also broadening, there are still certain areas that are relatively unexplored, including the formalization of pattern creation and identification methods, as well as pattern evaluation and work on antipatterns. Ontology pattern research is most often validated by way of theory, including through examples and feature comparisons.

The analysis presented indicates that some work remains to be done in formalizing:

- methods of evaluating the efficiency and effectiveness of ODPs,
- the development ODPs for particular usages,
- finding ODPs in existing ontologies or other information artifacts.

Applications of information logistics could benefit from the use of ontologies, and the encoding of information demand patterns into ontology patterns could help in transferring knowledge about information demand and distribution. There exists already a number of published patterns as well as notation and methods that can prove useful in such modeling work.

References

1. Tarasov, V., Lundqvist, M.: Modeling collaborative design competence with ontologies. *International Journal of e-Collaboration* **3**(4) (2007) 46–62

³ <http://ontologydesignpatterns.org/>

⁴ http://ontologydesignpatterns.org/wiki/Submissions:Information_realization

⁵ http://ontologydesignpatterns.org/wiki/Submissions:Time_indexed_participation

⁶ <http://ontologydesignpatterns.org/wiki/Submissions:AgentRole>

2. Sandkuhl, K., Billig, A.: Ontology-based artefact management in automotive electronics. *International Journal of Computer Integrated Manufacturing* **20**(7) (2007) 627–638
3. Smirnov, A., Pashkin, M., Chilov, N., Levashova, T.: Knowledge logistics in information grid environment. *Future Generation Computer Systems* **20**(1) (2004) 61–79
4. Blomqvist, E., Sandkuhl, K.: Patterns in ontology engineering: Classification of ontology patterns. In: *Proceedings of the 7th International Conference on Enterprise Information Systems*. (2005) 413–416
5. Gangemi, A.: Ontology design patterns for semantic web content. In: *The Semantic Web–ISWC 2005*, Springer (2005) 262–276
6. Presutti, V., Gangemi, A., David, S., de Cea, G., Surez-Figueroa, M., Montiel-Ponsoda, E., Poveda, M.: NeOn Deliverable D2. 5.1. A Library of Ontology Design Patterns: reusable solutions for collaborative design of networked ontologies. NeOn Project. <http://www.neon-project.org> (2008)
7. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: *Design patterns: elements of reusable object-oriented software*. Addison-wesley Reading, MA (1995)
8. Fowler, M.: *Analysis Patterns: reusable object models*. Addison-Wesley (1997)
9. Hay, D.: *Data Model Patterns: Conventions of Thought*. Dorset House (1996)
10. Fowler, M.: *Patterns of enterprise application architecture*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA (2002)
11. Buschmann, F., Henney, K., Schmidt, D.: *Pattern-oriented software architecture: On patterns and pattern languages*. John Wiley & Sons Inc (2007)
12. Hammar, K., Sandkuhl, K.: The state of ontology pattern research - a systematic review of iswc, eswc and aswc 2005-2009. In: *Proceedings Of The Workshop On Ontology Patterns (WOP 2010) At The 9th European Semantic Web Conference (ISWC 2010)*, Shanghai (November 2010) 1–13
13. Ivarsson, M., Gorschek, T.: Technology transfer decision support in requirements engineering research: a systematic review of REj. *Requirements engineering* **14**(3) (2009) 155–175
14. Sandkuhl, K.: Information logistics in networked organizations: Selected concepts and applications. In: *Enterprise Information Systems*. (2009)
15. Lundqvist, M., Sandkuhl, K., Seigerroth, U.: Modelling information demand in an enterprise context : Method, notation, and lessons learned. *International Journal of Information System Modeling and Design (IJSMD)* **2**(3) (2011) ?–?