A Survey on Domain Knowledge Representation with Frames

Vladislavs Nazaruks and Jānis Osis
Department of Applied Computer Science, Riga Technical University, Sētas iela 1, LV-1048, Riga, Latvia

Keywords: System Modelling, Frames, Knowledge Base.

Abstract: Domain knowledge acquisition, presentation and maintenance play an important role in software development. Frame-based knowledge bases are used to support the decision-making process. We believe that a use of a knowledge base that supports model transformations is not less important. To clarify the current state of a use of frame systems we have investigated recent research in the field to find out about techniques used for knowledge acquisition, frame elements, implementation technologies, existing limitations in implementation and integration with other knowledge representation formats. The overview showed that knowledge acquisition often is manual, procedural knowledge in frames can be separated, web-enabled knowledge bases are the trend, and the frame systems can be used in hybrid knowledge bases. However, some limitations in performance and integration with other knowledge representation systems exist due to support of different world paradigms. The obtained results show that despite existing limitations, frame-based knowledge systems still are in use and researchers found ways how to adapt them to the modern requirements.

1 INTRODUCTION

One of unsolved issues in software development is inconsistency between the problem and its solution. Some of the causes are insufficient understanding of characteristics of the problem domain, their instability, and a lack of formal means to represent, maintain and provide domain knowledge for all the stakeholders in the development process.

There are three core things that could be applied to avoid (or solve) this issue, namely, unification, standardisation and solid theories (Osis & Asnina 2011; Osis & Asnina 2014). Unification lead up to the origin of Unified Modelling Language (UML) and Business Process Model and Notation (BPMN). Standardisation is an ongoing process, where different technologies and approaches obtain common interfaces that allow their common usage and integration. Development of the solid theory (or theories) is the hardest one.

The solid theory requires accurate understanding of domain phenomena. The appearance of Model Driven Architecture (MDA) (Miller & Mukerji 2001) can be seen as the first step. MDA can be defined as “an approach to system development and interoperability that uses models to express and direct the course of understanding, requirements elicitation, design, construction, deployment, operation, maintenance, and modification” (OMG 2010). The key principle of the MDA is a separation of concerns in specifications. Thus, a system is analysed from three viewpoints and corresponding models, namely a computation independent model (CIM), a platform independent model (PIM) and a platform specific model (PSM).

The CIM specifies domain knowledge, business rules and processes, data vocabulary, and requirements (Miller & Mukerji 2001). The CIM can be represented by means of BPMN (Rhazali et al. 2016), Data Flow Diagrams (Kardoš & Drozdová 2010), UML diagrams (e.g. use case and activity diagrams), user stories, business rules (Essebaa & Chantit 2016), etc. Often it is unstructured text, structured text, or semi-formal (or even informal) graphical representations. These formats cannot keep all the knowledge about the domain in the format understandable for automated processing and inferring. Therefore, one of the questions is how to represent domain knowledge to have enough facts and rules on the domain to automate knowledge processing for model to model transformations.
We believe that a domain knowledge base (or bases) can serve as storage of this knowledge in the consistent and computer-understandable format (Figure 1). The knowledge base inferring mechanism must help in avoiding untrue facts about the domain. The knowledge base should maintain domain knowledge related to system functionality, behaviour and structure. Then this knowledge can be used in the further analysis models that should be obtained by automated transformation of this knowledge model. Knowledge acquired from the solution domain must be in consistency with problem domain knowledge. Thus, consistency between corresponding analysis models also should be kept. Then this analysis models could be refined.

There are several formats for knowledge representation such as frame networks, ontology, concept networks, product rules etc. The very interesting idea of keeping natural language semantics in text in the so-called artificial natural languages (e.g., Esperanto, Conlang, Lingvata) is expressed in (Roux 2013), where the author states that this may help translate semantics correctly from one language to another. However, this field is very specific and is out of scope of this paper.

Each of knowledge representation formats has its own advantages and limitations. The limitations are caused by two of the most fundamental problems in the field of expert systems, namely knowledge acquisition and representation (Kornienko et al. 2015) and search (Kornienko et al. 2015; Marinov 2004).

As Minsky defines in (Minsky 1974), “a frame is a data structure for representing a stereotyped situation”. A frame can be thought of as a network of nodes and relations. The “top levels” of a frame are fixed, and represent things that are always true about the supposed situation. The lower levels have many terminals — “slots” that must be filled by specific instances or data.

Collections of related frames are linked together into frame systems. Originally, for implementation of frame systems a frame representation language (Foster & Juell 2006) was developed on the base of Lisp programming language, however it did not continue to be maintained and repaired. At the present, there is no single language for implementation of frames.

The distinction of the frame systems and ontologies is in different inferring mechanisms. The frame system supports the so-called closed-world inferring paradigm (Detwiler et al. 2016), where all facts that are presented in the system are true. If some fact is not presented, that means that it is untrue. It allows avoiding errors in inferring mechanism related to the knowledge representation format. Other formats as, for example, ontology, may support the open-world paradigm, where all facts that are not presented may also be true. This can lead to errors in inferring mechanism, but on the other hand may allow an engineer to discover new knowledge from the existing facts. A frame-based knowledge base is one of the typical models or a part of such models for knowledge representation in expert and decision-making systems (Kornienko et al. 2015).

In the current research, we overview research work that deal with frame systems. The goal of this
review is to summarize information about knowledge acquisition, frame elements, implementation technologies, existing limitations in implementation and integration with other knowledge representation formats.

Section 2 states the research questions and research method. Section 3 gives an overview of the related research. Section 4 summarizes the results.

2 PROCESS OF THE REVIEW

As previously mentioned, the knowledge base for maintaining declarative and procedural knowledge about the domain should be suitable for automated model transformations. The frame systems can be well integrated with the object-oriented paradigm used in the MDA models. Therefore, we would like to get answers of the following questions:

- Q1. What is a way of entering knowledge into the frame systems: manual or automated?
- Q2. What are domains of knowledge represented in the frame systems?
- Q3. What technologies are used to implement frame systems?
- Q4. What technologies are used to implement frame systems?
- Q5. What limitations exist?
- Q6. Does integration with other knowledge representation systems exist, and what they are?

The research procedure was the following. We have investigated the following publication databases IEEE Xplore Digital Library, SpringerLink, ScienceDirect, and ACM Digital Library. The search keywords used were “frame-based knowledge system”, “frame-based knowledge base”, “frame system”, “knowledge frame”, “knowledge base”.

In the ACM Digital Library, there were found 14 publications on frames, 4 of whom were excluded since either were not related to the frames, or contained only detailed information of frame elements. The following 10 publications on frames were overviewed: (Beltrán-Ferruz et al. 2004; Corcoglioniti et al. 2016; Foster & Juell 2006; Gennari et al. 2005; Grigorova & Nikolov 2007; Kramer & Kaindl 2004; Marinov 2004; Rector 2013; Kim et al. 2008; Tan et al. 2007).

In the ScienceDirect database, there were found the following 8 publications: (Al-Saqqar et al. 2016; Binba et al. 2016; Detwiler et al. 2016; Hernández & Serrano 2001; Marinov 2008; Shiue et al. 2008; Xue et al. 2012; Zopounidis et al. 1997).

In the SpringerLink database and the IEEE Xplore Digital Library, there were found 2 publications: (Tettamanzi 2006) and (Xue et al. 2010), correspondingly.

3 DOMAIN KNOWLEDGE REPRESENTATION WITH FRAME SYSTEMS

The selected publications are overviewed in the chronological sequence.

Frames as a part of hybrid knowledge-based system are presented in (Hernández & Serrano 2001), where they are integrated with product rules, constraints and other knowledge representation techniques. The knowledge base represents control knowledge and domain knowledge required for the emergency system. Design and implementation are done using Knowledge Structure Manager tool and the corresponding methodology.

The manual knowledge acquisition is described in (Beltrán-Ferruz et al. 2004), where the frame system represents Mikrokosmos system that is used in knowledge-based machine translations. The Mikrokosmos contains definitions of concepts that correspond to classes of things and events in the world. The world model is organized as objects, events and properties set in a complex hierarchy. The format of Mikrokosmos Ontology formally introduces its syntax and semantics using a BNF grammar. Ontology is saved in a text file using Spencer notation that is based on XML. Another possibility is to use notation called Beale that is based on Lisp. The reasoning is implemented using JENA and the DIG interface, as well there are two different inference engines: RACER and FaCT. For ontology implementation, they have developed an import plugin for Protege 2.0. The frames used contain a concept name, slots, corresponding facets and filler(s). There are two kinds of slot fillers: (1) ATTRIBUTE or RELATION, that represent links between concepts in the hierarchy; (2) special ONTOLOGY-SLOTS dedicated to determining the structure of ontology. Possible descendants for the latter one are DEFINITION, DOMAIN, INSTANCES, INVERSE, IS-A, RANGE and SUBCLASSES. A filler generally contains either a name of a concept of the ontology or an instance. The authors mention that some special slots limit expressiveness. In the research, the authors provide mappings from Mikrokosmos Ontology frame system to descriptive logic language OWL that supports SHIQ logic. Because of the transformation, some knowledge is kept as annotations to description.
logic concepts due to the limitation of description logic expressiveness.

The authors in (Kramer & Kaindl 2004) also discuss knowledge frame-based systems, which contain frames with slots with values and rules in form IF-THEN that are not attached to frames. The authors believe that evaluation of modularization metrics such as coupling and cohesion can help in assessment of technical quality of such systems that may suffer from inadequate representation of knowledge. This evaluation should help quickly understand potential problems with constructs in the knowledge base.

The implementation of frames using XML (eXtensible Markup Language) is discussed in (Marinov 2004). Evolution of web-based knowledge modelling languages opens new opportunities for the application of Web-oriented frame-based knowledge representation. The new generation of such languages is DAMPL (DARPA Agent Markup Language), OIL (Ontology Inference Layer) and SHOE (Simple HTML Ontology Extension). They are system independent and web compatible thanks to XML and RDF support. The frame structure considered in the research is classical one, where a frame consists of slots, facets, and active slots with query procedures and daemons, which represent product rules. As the author mention in (Marinov 2008), tools for implementation of frame systems may be Protégé 2000, Conceptually Oriented Design/Description Environment (CODE4), and FrameD (a distributed object-oriented database).

The authors in (Gennari et al. 2005) investigate transformations between two fundamentally different knowledge representation formats, namely frame systems and relational databases. A frame-based system was developed in Protégé that subscribes to the Open Knowledge Base Connectivity. The authors declare that a frame-based system has greater expressiveness, and such transformation should lead to the loss of some information.

The author in (Tettamanzi 2006) proposes a frame-based formalism for representing imprecise knowledge, combining it with fuzzy logic. The formalism is based on frames, but frames are simplified. According to this formalism, knowledge consists of three basic types of objects: (1) knowledge elements, which can be either atomic (atoms) or complex (frames); (2) fuzzy sets, or linguistic values; and (3) relations, which can be fuzzy rules or subsumption relations.

The list of the basic methods for knowledge representation (Grigorova & Nikolov 2007) state that classification and inheritance in frame systems support knowledge engineering efficiently. The authors mean three disadvantages of frame-based knowledge base: (1) it should work with completely known object characteristics; (2) the knowledge domain has to be static; and (3) the fact that procedural knowledge is represented with programming code inside the frame does not allow reasoning about this knowledge (reasoning could be done with it). The authors declare that frames are a proper method for knowledge representation when the goal is realization of a natural language sentences analyser.

In (Tan et al. 2007), the authors use ontologies as a structured and semantic representation of domain knowledge. The authors propose a method for building frame-based corpus for the domain of biomedicine based on domain knowledge provided by ontologies. They compared one frame to the corresponding frame in BioFrameNet, and examined the gaps between the semantic classification of the target words in the domain-specific corpus and in FrameNet and PropBank/VerbNet.

The authors in (Kim et al. 2008) describe frame application for probabilistic dialog systems, namely, they have introduced a frame-based state representation. The frames have slots that can dynamically update their values, and frames can be grouped per indistinguishable user goal states. Knowledge accusation methods and integration with other knowledge representation formats are not discussed.

Application of frames in banking expert systems, which requires fast and up-to-date decision-making is discussed in (Shiue et al. 2008). The basic knowledge representation format for such systems are rules that in the long run lead to decrease of the understandability and accessibility of the knowledge base as well as to increase of the complexity of maintenance of knowledge rules. The knowledge acquisition is manual. The frames are chosen since they provide “an easy way for encapsulating declarative knowledge with procedural one”. The frames contain slots with decision-making criteria, and corresponding procedures. Frame system maintenance foresees changing rules stored in objects. To create web-enabled interface for the expert system, the authors use Jess (Java Expert System Shell) and Object Web model, while an UML class diagram is used for representation of the model of the frame system. JavaBeans structures encapsulate the knowledge objects in the knowledge base. The limitation of such implementation is decrease of the performance of the system inference
and execution, when the knowledge frame structure gets more complicated.

The authors in (Xue et al. 2010) propose a frame-based ontological view specification language (FOSL) that is based on the knowledge frame paradigm and uses XML as encoding. XML documents that hold knowledge frames allow development of web-enabled information systems. The authors suggest using the “ontological views” on the conceptualization of the domain, thus establishing a possibility of several such views, and integration of ontological languages. The language uses a concept of a frame that has four standard levels: frame, slot, facet, and data. The authors also indicate on frame paradigm characteristics, i.e. a trade-off between the good expressiveness and the ease of inference. Continuing their research in (Xue et al. 2012), the authors analyse the capabilities of semantic integration on the basis of the frame-based ontological view, which can be created from the information model of an information system.

The interesting discussion about representation and inference possibilities of ontologies, frames and UML is given in (Rector 2013). The author distinguishes knowledge representation systems into template-based (frames and UML) and axiom-based (ontology) by their inference mechanisms. It considers classical frame structures provided by Protégé-Frames.

The authors in (Sim & Brouse 2014) uses Protégé-Frames tool for implementation of OntoPersonaURM model that consists of three interrelated ontologies. This system like a system in (Shiue et al. 2008) is foreseen to be web-enabled and able to check constraints and run queries on the ontology by using PAL plug-in toolset of Protégé-Frames.

The authors in (Al-Saqqar et al. 2016) use knowledge frames to represent agent knowledge and correspondent communicative commitments presented by modal logic in multi-agent systems. The knowledge frame based system is integrated with model checking mechanism. However, it is not clear what method is used to acquire knowledge to the frame system.

The authors in (Bimba et al. 2016) provide very exhaustive survey on knowledge representation, implementation and acquisition techniques (for 2000–2015 years), i.e. the linguistic knowledge bases, expert knowledge bases, ontology and cognitive knowledge bases. The authors stress that a use of production rules that contain expert knowledge is not suitable for every knowledge type. The authors found out that knowledge acquisition is manual through communication with domain experts with some automated methods such as LSPE and acquiring English sentences from the Open Mind Common Sense (OMCS) corpus. The authors have summarized that there are different applications for Natural Language Processing, Question Answering, Information extraction/retrieval, classification, knowledge discovery, engineering, health care, education, finance, environment, business, machine learning, robotics and forecasting. To implement linguistic knowledge bases, FrameNet, WordNet and ConceptNet supporting tools are used such as Sesame, SWI, NTLK, and ADW. Programming and mark-up languages used in such implementations may be XML, Python, Java, SQL, RDF, SPARQL, Perl, and JSON. In case of product rules in expert systems Prolog is also used. In most linguistic knowledge bases, frame elements, semantic networks and semantic graphs (of frames, lexical semantic associations between synsets and graphs) are used, while expert knowledge bases make a use of IF-THEN rules that join linguistic objects, values and operators. One of big limitations of linguistic knowledge bases are their dependence on volatile expert knowledge, and expensiveness and difficulty in building and expending the base. Besides that, frame nets cannot handle text coherence and link arguments across sentences. As the authors mention, FrameNets integration with other knowledge representation techniques is very difficult, and transformation to them requires additional effort in preserving richness of its annotations.

The authors in (Corcoglioniti et al. 2016) suggest a new approach called PIKES that is implemented as an open-source Java application. This approach is dedicated to analysis of any text in a natural language by means of several Natural Language Processing tools. The resulting net is implemented in RDF (Resource Description Framework) knowledge graph by means of extended RDFpro tool and support of SPARQL-like rule evaluation, where instances are linked to matching entities in DBpedia using OWL elements, and typed according to classes encoding VerbNet, FrameNet, PropBank and NomBank frame types.

The authors in (Detwiler et al. 2016) describe the transformation of the Foundational Model of Anatomy (FMA) represented in a frame system to the modern semantic web language OWL2. The main difficulty of the transformation lies in the difference between closed-world assumption made in frame systems and open-world assumption made in the ontology. The open-world assumptions may lead to mismatches in interpretations, since from its point of view even if some fact is not stated in the system, it
does not mean that this fact is untrue (as it is in frame systems). Therefore, some untrue facts may be considered as possible. The frames in FMA are characterized by names, slots and facets (that are used as constraints to slot values). But there is no straight correspondence between slots and class properties in OWL, since slots may represent as frame structural properties as relations to other frames. Besides that, in frames the semantic of slots without values is not clearly defined. Although it is possible to have such slots in frames, in the ontology only those properties that must have values are to be listed.

4 DISCUSSION AND CONCLUSIONS

Summarizing overview results here and in tables (Table 1, Table 2, Table 3), it is possible to conclude that the most common way of entering knowledge into the frame system is manual, i.e. a knowledge engineer enters facts and assertions about the domain based on results of interviews with domain experts and other information about the domain. Only in case of text analysers automated entering is applied, but the amount of human participation in this process is not clear.

The frame-based representation of declarative and procedural knowledge has a wide application, but the last decade tendency is the health care and biomedicine (mostly for ontologies of terms), forecasting and text/natural language processing (Q2 in Table 1).

The most used implementation tool is Protégé-Frames (Table 2). The implementation languages differ from the specific knowledge representation languages like FOSL to general ones like Java. The overview showed that many frame-based knowledge systems are often integrated with other ontology nets. Besides that, there is a tendency to make frame systems more web oriented (Table 2).

Limitations mentioned by authors are inadequate representation of knowledge (Kramer & Kaindl 2004), greater expressiveness that can lead pure ontologies to the loss of information in case of transformation into them (Gennari et al. 2005; Bimba et al. 2016; Detwiler et al. 2016), necessity to work with the completely known characteristics and static knowledge domain (Grigorova & Nikolov 2007), representation of the procedural knowledge as programming code inside frames (Grigorova & Nikolov 2007), and the fact that complex structures can decrease the performance of the system inference and execution (Shiue et al. 2008; Xue et al. 2010).

There could be integration with other knowledge representation systems such as product rules and business constraints (Hernández & Serrano 2001), OWL (Hernández & Serrano 2001; Corcoglioniti et al. 2016; Detwiler et al. 2016), fuzzy logic (Tettamanzi 2006), and modal logic (Al-Saqqar et al. 2016).

The obtained results show that frame systems are still in use. There are made optimistic attempts to adapt this knowledge representation format to new technologies, especially web technologies. This

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. What is a way of entering knowledge into the frame systems: manual or automated?</td>
<td>(Beltrán-Ferruz et al. 2004; Shiue et al. 2008; Bimba et al. 2016; Detwiler et al. 2016)</td>
</tr>
<tr>
<td>Manual</td>
<td></td>
</tr>
<tr>
<td>Automated</td>
<td>(Grigovora &amp; Nikolov 2007; Xue et al. 2010; Xue et al. 2012; Corcoglioniti et al. 2016)</td>
</tr>
<tr>
<td>Not discussed (assumed to be manual)</td>
<td>(Hernández &amp; Serrano 2001; Gennari et al. 2005; Tettamanzi 2006; Tan et al. 2007; Kim et al. 2008; Shiue et al. 2008; Rector 2013; Sim &amp; Brouse 2014; Al-Saqqar et al. 2016)</td>
</tr>
<tr>
<td>Q2. What are domains of knowledge represented in the frame systems?</td>
<td></td>
</tr>
<tr>
<td>Emergency systems</td>
<td>(Hernández &amp; Serrano 2001)</td>
</tr>
<tr>
<td>Machine translation</td>
<td>(Beltrán-Ferruz et al. 2004; Bimba et al. 2016)</td>
</tr>
<tr>
<td>Biomedicine, health care</td>
<td></td>
</tr>
<tr>
<td>Probabilistic dialog systems (forecasting)</td>
<td>(Tan et al. 2007; Detwiler et al. 2016; Bimba et al. 2016)</td>
</tr>
<tr>
<td>Banking expert systems</td>
<td></td>
</tr>
<tr>
<td>Not domain-specific (Natural language processing, question answering, information extraction/retrieval, classification, machine learning, robotics)</td>
<td>(Kramer &amp; Kaindl 2004; Marinov 2004; Marinov 2008; Gennari et al. 2005; Tettamanzi 2006; Grigorova &amp; Nikolov 2007; Xue et al. 2010; Xue et al. 2012; Rector 2013; Sim &amp; Brouse 2014; Bimba et al. 2016; Al-Saqqar et al. 2016)</td>
</tr>
</tbody>
</table>
allows integrating frame-based knowledge systems with already existing ontologies and other knowledge representation techniques. This means that frame systems can be applied also for our purpose considering enumerated limitations and possibilities.

The further research is related to the design and implementation of the frame-based or hybrid knowledge base for software development based on the model-driven paradigm. As mentioned in Introduction, the goal of this knowledge base is to support system modelling from the computation independent viewpoint and corresponding knowledge transferring to functioning, behavioural and structural design models. Based on this research result, it become clear that we should pay strong attention to this system flexibility, maintainability, performance and integration with web technologies.

**REFERENCES**


Roux, C., 2013. Can “Made Up” Languages Help Computers Translate Real Ones?. Available at:


