IMPLEMENTATION OF INTENTION-DRIVEN SEARCH PROCESSES BY SPARQL QUERIES

Olivier Corby
INRIA Sophia Antipolis, BP 93, FR-06902 Sophia Antipolis, Cedex France

Catherine Faron-Zucker
I3S, Université de Nice Sophia, CNRS, BP 121, FR-06903 Sophia Antipolis, Cedex France

Isabelle Mirbel
INRIA Sophia Antipolis and I3S, Université de Nice Sophia, CNRS, France

Keywords: Search Process Modeling, Process Guidance, Ontology, Semantic Web, SPARQL, RDF(S).

Abstract: Capitalisation of search processes becomes a real challenge in many domains. By search process, we mean a sequence of queries that enables a community member to find comprehensive and accurate information by composing results from different information sources. In this paper we propose an intentional model based on semantic Web technologies and models and aiming both at the capitalization, reuse and sharing of queries into a community and at the organization of queries into formalized search processes. It is intended to support knowledge transfer on information searches between expert and novice members inside a community. Intention-driven search processes are represented by RDF datasets and operationalized by rules represented by SPARQL queries and applied in backward chaining by using the CORESE semantic engine.

1 INTRODUCTION

The study of research projects aiming at supporting activities of community members through a collective semantic memory (Ait Ameur and al., 2008; Yurchyshyna et al., 2008) highlights the need for query capitalization. Beyond *query* capitalisation, the capitalisation of whole *search processes* becomes a real challenge in many domains. By search process, we mean a sequence of queries enabling to find comprehensive and accurate information by composing results from different information sources. These processes are often difficult to acquire by novice users and they become more and more critical because of the specialization and proliferation of knowledge sources (Bhavnani et al., 2003).

(Bhavnani et al., 2003) presents an approach to formalize critical search procedures in the medical domain into strategy hubs. A search procedure is represented by an ordered set of sub-goals and strategy hubs provide search procedures and associated high-quality links to information sources that enable users to find comprehensive and accurate information. (Buffereau et al., 2003) propose to support navigation among resources by e-road maps which are composed of *steps* characterized by an *intention* or title, an optional subject and optional illustrations (web ressources). In this paper we propose a model to capitalize, reuse and share search queries and to organize them into formalized search processes. The originality of our proposal with regards to both strategy hubs and e-road maps relies in that we statically associate resource patterns to steps of search processes. Resources are dynamically selected when rendering queries associated to search process sub-goals.

Our paper is organized as follows. Section 2 presents the concept of semantic search process and the model we propose. Section 3 presents how search processes are operationalized by rules. We conclude in Section 4.

2 SEMANTIC SEARCH PROCESS

We define the notion of search process as a sequence of atomic searches to be processed by a domain expert to fulfill a domain-specific task or pro-

Corby O., Faron-Zucker C. and Mirbel I. (2009).
IMPLEMENTATION OF INTENTION-DRIVEN SEARCH PROCESSES BY SPARQL QUERIES.
In *Proceedings of the 11th International Conference on Enterprise Information Systems - Artificial Intelligence and Decision Support Systems*, pages 339-342
DOI: 10.5220/0001993103390342

cess. A search process may be seen as a particular kind of business process limited to search activities. Different business process modeling formalisms have been proposed in the literature. They can be classified into three categories: activity-oriented, product-oriented and decision-oriented ones (Nurcan and Edme, 2005). Decision-oriented models are semantically more powerful than the two others because they explain not only how the process proceeds but also why. Their enactment guide the decision making process that shapes the process, and helps reasoning about the rationale (Nurcan and Edme, 2005).

To support knowledge transfer about search process from experts to novices, we are concerned with the modeling of why the search process is decomposed the way it is, as well as with the specification of how it is decomposed. Moreover, to handle different users'profiles and levels of knowledge, we want to provide means to specify search processes at different levels of detail. For all these reasons, the approach we propose to model search processes is based on the adaptation of an intentional process modeling formalism (Rolland, 2007).

2.1 Maps

According to (Rolland, 2007), a map is a process model in which a non-deterministic ordering of intentions and strategies has been included. In our case, we focus on search intentions and search strategies. A map is a labeled directed graph with intentions as nodes and strategies as edges between intentions. A search intention is a goal that can be achieved by following a search strategy. An intention expresses what is wanted, a state or a result that is expected to be reached disregarding considerations about who, when and where. There are two distinct intentions that represent the intentions to start and to stop the process. A map consists of a number of sections each of which is a triple (source intention, target intention, strategy). A strategy characterizes the flow from the source intention to the target intention and the way the target intention can be achieved. A map contains a finite number of paths from its start intention to its stop intention, each of them prescribing a way to achieve the goal of the search process under consideration.

Let's take the example of a teacher belonging to a community of teachers and looking for resources to build a course about relational database. To fulfill this goal, s/he may search for resources about database history to build the introduction of its course, then resources about the relational model, normal forms and SQL, and finally resources on how to interact with a database from a programming language. *Search for resources about database history* and *search resources about relational model* are examples of intention. Depending on the context of the course (the number of teaching hours, the audience, ...), the teacher may be or not interested by some resources, for instance resources about database history. Consequently, this intention looks optional and two paths are suggested to fulfill the main goal of building a course about relational database with or without the *Search for resources about database history* intention. The formalization of this scenario with the map model is presented in figure 1(a).

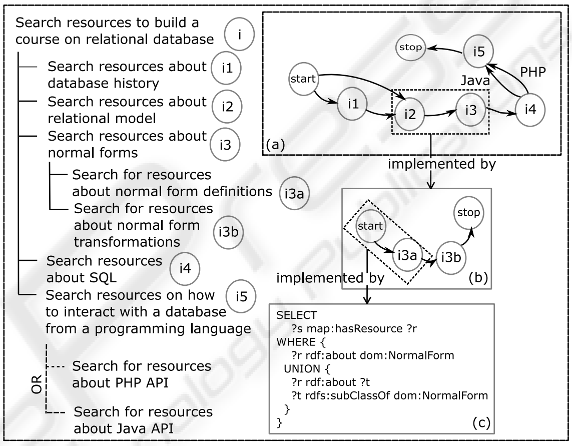

Figure 1: Example of a Map.

As shown in figure 1(a), there might be several flows from a source intention to a target intention, each corresponding to a specific strategy. It is the case of intention `i5` labeled by *search for resources on how to interact with a database from a programming language* which is reachable via a strategy based on the Java programming language or via a strategy based on the PHP programming language. There might also be several strategies to reach a target intention from different source intentions. It is the case of intention `i2` labeled by *search for resources about the relational model* which is reachable from intention `i1` labeled by *search for resources about database history* or directly from the start intention.

Still according to (Rolland, 2007), the execution of each section of a map is supported by an intention achievement guideline (IAG) which provides an operational or an intentional means to fulfill the target intention. In our work, we operationalize an IAG by the execution of a query on the community semantic memory or we define it by a refined map, as shown in figure 1(b) and (c).

To further formalize intentions and strategies, we rely on (Prat, 1999) proposal, which has already proven to be useful to formalize goals (Ralyte, 2001;

Rolland, 2007). According to (Prat, 1999), an intention statement is characterized by a verb and some parameters which play specific roles with respect to the verb. Among the parameters, there is the object on which the action described by the verb is processed. In (Ralyte, 2001; Rolland, 2007) different other relevant parameters have been identified : beneficiary, reference, quality, quantity, direction, time, ways (manner or means) and location. Let us consider again the map depicted in figure 1(b). Intention `i3a` labeled by *search for resources about normal forms definition* is described by its verb *search* and its object *resources about normal forms definition*. Intention `i5` labeled by *search for resources on how to interact with a database from a programming language* achieved through the *PHP* strategy is described by its verb *search*, its object *API* and its manner *PHP*.

2.2 Search Process Modelling

Starting from the map model, we propose a search process modelling based on Semantic Web standards. We gathered the concepts and relationships of the map model and we built an RDFS ontology dedicated to the representation of search processes (including for instance classes *Verb*, *Object* and *Parameter* to represent intention statement).

In order to specify intention statements, we are currently exploiting a single verb, namely class *Search* which instantiates class *Verb*, and we consider many domain-specific concepts as instantiations of class *Object*. The class *Parameter* is instantiated into several classes modeling the context of the search process. We distinguish between domain dependent and domain independent contextual information. To further populate the search process ontology with classes of parameters, a domain independent *context ontology* is under development and mappings with concepts from a domain-specific ontology as well.

Based on the search process ontology, search processes (or fragments of search process) are then represented by RDF annotations.

By relying on RDF(S) which is now a widespread Web standard, we ensure the capitalization, reuse and share of these representations of search processes among community members. Beyond an alternative way to organize and to dynamically access resources in a community memory, we provide means to capitalize search processes themselves. We take advantage of the inference capabilities provided by the RDF framework to reason on search process representations, especially to organize them and retrieve them for reuse.

3 INTENTION ACHIEVEMENT GUIDELINE MODELLING

Reusing search processes (or search process fragments) is intended to enable a dynamic connection of different search processes and therefore the building of a whole search process by combining those (fragments of) search processes which both satisfy the global intention and retrieve available resources. This goes through modeling IAGs which connect a section of a map representing a search process either with another map representing another search process (intentional means) which can be viewed as a fragment of the global process fulfilling the target intention of the connected section, or directly with a query (operational means) retrieving relevant resources in the community semantic memory.

We propose to represent an IAG by a rule which conclusion represents a section of a map and which premise represents either an operational means (a query) or an intentional means (a map) fulfilling the target intention of the section in conclusion. We call a rule *concrete* or *abstract* depending on wether its premise represents operational or intentional means. The SPARQL language provides a unified framework to represent both concrete and abstract rules through the CONSTRUCT query form. A CONSTRUCT query form returns an RDF graph specified by a graph template in the query form and constructed by taking each query solution, substituting for the variables in the graph template and combining the resulting RDF triples. In our case we formalize a rule representing an IAG by a SPARQL query which CONSTRUCT clause is the conclusion of the rule, i.e. the graph template to construct the RDF representation of a section of a map (the conclusion of the rule) and which WHERE clause is the premise of the rule, i.e. a graph pattern representing a map (abstract rule) or criteria for retrieving relevant ressources (concrete rule).

In our running example, let us consider again the IAG associated to the section of the map presented in figure 1 aiming at searching for resources about normal forms. This IAG describes an intentional means to fulfill target intention `i3` of the section. We represent it by an abstract rule formalized by the following SPARQL query, where prefix `map` refers to the search process ontology and prefix `d` refers to a domain ontology on relational database:

```
CONSTRUCT {                                (1)
  _:s map:hasTarget _:i                    (2)
  _:i map:hasObject d:NormalForm           (3)
  _:s map:operationalizedBy ?g             (4)
}                                          (5)
WHERE {                                    (6)
```

```
 GRAPH ?g {                              (7)
  ?s1 map:hasSource ?i0                  (8)
  ?i0 rdf:type map:Start                 (9)
  ?s1 map:hasTarget ?i1                 (10)
  ?i1 map:hasObject d:NormalFormDef     (11)
  ?s2 map:hasSource ?i1                 (12)
  ?s2 map:hasTarget ?i2                 (13)
  ?i2 map:hasObject d:NormalFormTransform
  ?s3 map:hasSource ?i2                 (15)
  ?s3 map:hasTarget ?i4                 (16)
  ?i4 rdf:type map:Stop                 (17)
 }
}
```

The CONSTRUCT clause of this query is a graph template for building an RDF graph representing any section aiming at searching for resources about normal forms. It includes both statements (lines 2-3) describing the object of the target intention of the section with the domain concept *NormalForm* instantiating concept *Object* of the map ontology and a statement (line 4) about the RDF graph operationalizing the section and which content is described in the WHERE clause of the query. This links together the two levels of intention refinement.

The WHERE clause of the query describes how to operationalize any section (in particular the one of our example) whose RDF representation matches with the graph template in the CONSTRUCT clause. It is a graph template that matches with the RDF representation of the map shown in figure 1. It includes statements about three sections: the first ones (lines 8-11) describe a first section `?s1` which source intention is a start and which target intention has for object the definition of a normal form; the following ones (lines 12-14) describe a second section `?s2` which source intention is the target intention of the first section `?s1` and which target intention has for object the transformation of normal forms; the last ones describe a third section which source intention is the target intention of the second section `?s2` and which target intention is a stop.

Instantiating search processes, i.e. combining sub-processes into a global process is achieved by applying rules implementing IAGs in backward chaining. The problem of operationalizing the initial strategy provided to the backward chaining engine then boils down to operationalizing the sections described in the WHERE clause of the query. We rely on the CORESE[1] (Corby et al., 2006) semantic engine for both backward chaining on the knowledge base of SPARQL queries and matching whith the knowledge base of RDF annotations of domain resources.

[1] http://www-sop.inria.fr/edelweiss/software/corese/

4 CONCLUSIONS

In this paper, we proposed an approach relying on semantic Web technologies and models to capitalize, reuse and share search queries and search processes. By modeling search processes, our aim was to capture knowledge and best practices into series of structured search activities. Therefore, starting from an intention driven process modeling formalism, we proposed an ontology to annotate search processes and we operationalized guidelines associated to search processes fragments with rules implemented as SPARQL queries. As a result, instantiation of search processes is supported by backward chaining among the rule base and matching with the RDF dataset annotating the community resources.

REFERENCES

Ait Ameur, Y. and al. (2008). Semantic hubs for geographical projects. In *Semantic Metadata Management and Applications (SeMMA), workshop at ESWC*.

Bhavnani, S., Bichakjian, C., Johnson, T., Little, R., Peck, F., Schwartz, J., and Strecher, V. (2003). Strategy hubs: Next-generation domain protals with search procedures. In *ACM Conference on Human Factors in Computing Systems*.

Buffereau, B., Duchet, P., and Picouet, P. (2003). Generating guided tours to facilitate learning from a set of indexed resources. In *IEEE International Conference on Advanced Learning Technologies (ICALT)*, Athens, Greece.

Corby, O., Dieng-Kuntz, R., Faron-Zucker, C., and Gandon, F. (2006). Searching the semantic web: Approximate query processing based on ontologies. *IEEE Intelligent Systems Journal*, 21(1).

Nurcan, S. and Edme, M. (2005). Intention-driven modeling for flexible workflow applications. *Software Process: Improvement and Practice*, 10(4):363–377.

Prat, N. (1999). *Réutilisation de la trace par apprentissage dans un environnement pour l'ingénierie des processus*. PhD thesis, Université Paris I - Sorbonne.

Ralyte, J. (2001). *Ingénierie des méthodes à base de composants*. PhD thesis, Université Paris I - Sorbonne.

Rolland, C. (2007). *Conceptual Modelling in Information Systems Engineering*, chapter Capturing System Intentionality with Maps. Springer-Verlag.

Yurchyshyna, A., Faron-Zucker, C., Mirbel, I., Sall, B., Le Thanh, N., and Zarli, A. (2008). Une approche ontologique pour formaliser la connaissance experte dans le modèle du contrôle de conformité en construction. In *19ième journées francophones d'ingénierie des connaissances*, Nancy, France.