ONTOLOGY-BASED APPROACH TO SEMANTIC QUERY OPTIMIZATION

Lela Tsitashvili[1], Badri Meparishvili[2]
1. Akhaltsikhe State Educational University,
2. Georgian Technical University

Summary

This paper is consecrated to a short review of query processing in heterogeneous models database systems and also a new approach to semantic search. In informational and database systems, web databases or internet, and other systems, data is collected from different information sources or generated by different users. Traditional approaches to database management have not paid much attention to the semantics of user queries. Query optimization uses rewriting techniques for reordering the operators in the relational algebra expressions and selecting a good implementation for each of them. The technique, which is based on semantic transformations to generate semantic equivalent queries, is called *semantic query optimization* (SQO). One of the classes of SQO can be represented the *ontological-based* approaches. This paper is consecrated to the techniques used for semantic query processing based on ontology. To match a relational database with an ontology we propose a matching process which includes a sequence of syntactic and semantic matching operations. We present a new approach to evolutionary optimization of the multi-agent synergetic model.

Keywords: semantic query. Ontology. Query optimization. Entropy.

1. Introduction

In informational and database systems, web databases or internet, and other systems, data is collected from different information sources or generated by different users. Query processing is one of the main task of a database management system. The queries are usually specified by imposing conditions on the attribute values stored in the database. Traditional approaches to database management have not paid much attention to the semantics of user queries. Query optimization uses rewriting techniques for reordering the operators in the relational algebra expressions and selecting a good implementation for each of them. The technique, which is based on semantic transformations to generate semantic equivalent queries, is called *semantic query optimization* (SQO).

The basic idea of semantic query optimization (SQO) is to reformulate a query to another more efficient query, which is semantically equivalent, i.e. provides the same answer [1,2,3,4]. SQO approaches use semantic knowledge in various forms including semantic rules and range rules. Range rules states facts about the range of values of a given attribute, whereas semantic rules define the regularity of data for a given database. Therefore, these rules can be driven from the nonuniform distribution of values in a database. Expressing semantics in the form of horn clause sets allows the optimizer to make possible reformulations on an input query involving the insertion of new literals, or the deletion of literals, or the refuting the entire query. Several approaches on SQO have been developed to address different aspects of query processing: In [5,6] semantic rules have been used to derive useful information, which can reduce the cost of query plans.

2. Use of Ontologies for Semantic Query Processing.

A classification for SQO approaches, based on the semantic knowledge, is presented into three classes: *static constraints-based*, *dynamic constraints-based*, and *ontological-based* approaches [7]. In database management systems, ontologies can play the role of a query model for processing user queries. Here, we propose an ontology-based semantic query processing approach for database systems. Ontologies can be represented by graph-based formalism. Its basic formal settings and some related operations are necessary for identifying concepts and relationships in an ontology. The mappings between Ontologies and Database schemas determine how concepts and relationships in an ontology are related to relations and attributes in the underlying database schema. The correspondences between ontology elements and database extension means how ontology elements (concepts and relationships) correspond to database instances (tuples, single values). Mapping can be defined as a graph of *mapping elements.* The main idea is to transform a semantic query before any query optimization. The system mainly consists of the transformation engine, which performs: a *syntactic transformation* to rewrite an input query into a canonical form, and a *semantic transformation* to produce another query that better meet user's intention using semantic rules [9]. These rules use semantic information extracted from an ontology.

3. Synergy-based Semantic Model

We model the problem of ontology matching as an optimization problem of a mapping between ontologies, and every ontology has its associated feature sets. As a powerful heuristic search strategy, evolutionary algorithms can be employed for the ontology matching problem. As the tools of mathematic formalism, we present a new approach which is based on the use of so called “a synergetic graph”, which represents the topological model of any complex system i.e. axon-dendrite model with synaptic connections between them [10,11,12].

Formally ontology model can be represented as the set of neurons in the form of a graph:

$$C = \{c_i\}, \quad i = \overline{1, N} \tag{1}$$

where: concept c_i can be represented as a set of properties:

$$c_i = \{p_{ik}\}, \quad k = \overline{1, L} \tag{2}$$

Generally, each of properties (or attributes) is defined as a tuple:

$$p_{ik} = \{s_{ik}, d_{ik}, \omega_{ik}\} \tag{3}$$

where: $s_{ik} \in \{-1, +1\}$ - is the sign of attribute;

$d_{ik} \in D$ - is the type of attribute;

$\omega_{ik} \in [0,1]$ - is the weight coefficient of attribute.

The total number of concepts: $Q = \sum_{i=1}^{N} \sum_{k=1}^{L} p_{ik}$ (4)

Interaction or connection between concepts is realized by the relations

$$R_{ij} = \{p_{ik} \circ p_{kj}\} \tag{5}$$

where: $\circ$ - is the relation or cohesion.

Each relations established in case of respective conditions:

$$R_{ij} = \left\{ (s_{ik} = -s_{kj}) \wedge (d_{ik} = d_{jk}) \wedge (|\omega_{ik} - \omega_{jk}| = \min_k \right\} \tag{6}$$

where: $s_{ik} = -s_{kj}$ is opposite polarity of attributes;

$d_{ik} = d_{jk}$ is the identity of types;

$(\omega_{ik} - \omega_{jk}) = \min$ is minimum difference of weight coefficients, which actually determines the degree of incompatibility.

Let us consider the environment as virtual element of system:

The weight coefficients for all its terminals will be $\omega_{or} = 0$, where: $r = \overline{1, F}$ and

A – number of relations: $A = Card\{R_{ij}\}$;

F –number of free attributes: $F = Q - 2A$;

μ – degree of incompatibility.

After redesignate $(ik) \to r_{(ik)}$, it will be: $\mu_{r_{(ik)}} = |\omega_{ik} - \omega_{jk}|$. (7)

Entropy is determined by the number of ways you could achieve a state. And entropy is calculated as the following function:

$$H = -\sum_{r_{(ik)}=1}^{F} \mu_{r(ik)} \log \mu_{r_{(ik)}} - \sum_{r_{(ik)}=1}^{R} P_{r_{(ik)}} (\mu_{r_{(ik)}} \log \mu_{r_{(ik)}} + (1 - \mu_{r_{(ik)}}) \log(1 - \mu_{r_{(ik)}})) \tag{8}$$

Ontology entropy is determined in the area of external and internal freedom. Compatibility of the relations is the necessary condition of ontology graph. Synergy, on its part, is the function:

$$S = \log \sum_{i=1}^{n} \mu_i - \sum_{i=1}^{h} p_i \log p_i \tag{9}$$

where: h – number of the orbits of isomorphic groups;

p – probability of the orbits of isomorphic groups.

4. Optimization

The main principle and originality of this approached consist in ontology model reconfiguration step-by-step for its improvement. We have not yet defined what it means to optimize of the graph by reconfiguration. After randomly formation of initial model, thought the instrumentality of the concept connectivity recombination is used to generate new solutions by criterion of the ontology entropy minimization or fitness function maximization.

Algorithmic Outline. In this section we present an algorithmic description. The research was focused on the new approach of ontology modeling and optimization based on structural transmutation of synergetic graph. Assuming that the set of concepts, the set of their attributes and with their signs, types and weight coefficients, limits of the weight coefficients and others constraints are given initial part of algorithm flow block proceeds as follow steps:

Step 1. Randomly formation of initial synergetic graph configuration (connectivity matrix) by each relation compatibility condition: $R_{ij} = \{(s_{ik} = -s_{kj}) \wedge (d_{ik} = d_{jk})\}$.

Step 2. Computing for each concept each relation degree of incompatibility μ by the equation: $\mu_{r_{(ik)}} = | \omega_{ik} - \omega_{jk} |$, entropy by the equation (expr. 8) and the entropy for initial ontology graph.

Step 3. There are two ways:

a) if graph entropy equals zero then go to step 12.

b) otherwise:

Step 4. Computing for gradient way to optimization.

Step 5. Transition in new point of gradient way.

Step 6. Defining of neuron behavioral vectors.

Step 7. Descending ranking of behavioral vectors by entropy.

Step 8. Internal cycle:

Correction of each vectors (entropy reduction) or graph reconfiguration.

Step 9. There are two ways:

a) if number of vectors equals zero then go to step 10.

b) otherwise return in step 8.

Step 10. Computing for current entire graph entropy.

Step 11. Verifying the condition of entropy decreasing.

There are two ways:

a) if entropy is decreasing then go to step 5.

b) otherwise:

Step 12. End of algorithm.

5. Conclusions

The main idea is to transform a semantic query before any query optimization. Query optimization uses rewriting techniques for reordering the operators in the relational algebra expressions and selecting a good implementation for each of them. Nowadays, conventional database querying does not always provide answers to users, which fully meet their expectations. One of the reasons, is that the query is treated at only the syntactical level. The technique, which is based on semantic transformations to generate semantic equivalent queries, called *semantic query optimization,* reduces to mappings between ontologies and database schemas.

In this paper, we have presented an evolutionary algorithm based approach for solving the ontology matching problem. When the semantic query is incoming, the matching can be determined through query and knowledge fragment graph intersection. In each case the realization of the following versions of optimization is possible by criterion of the entropy minimization. If this matching is comprehensive, then we have the case of disambiguation or ontologies similarity. Otherwise, it will be uncertainty, which depends on the value of matching entropy or graph difference.

References:

1. Fayyad, U, Piatetsky-Shapiro, G and Smyth, P, (1996), From data mining to knowledge discovery: an overview. In Fayyad, U, Piatetsky-Shapiro, G, Smyth, P and Uthurusamy, R (eds) *Advances in Knowledge Discovery and Data Mining*. AAAI Press, pp. 1–34.

2. Frawley, W, Piatesky-Shapiro,G and Matheus, C, (1991), Knowledge discovery in databases: an overview. AAAI/MIT Press, pp. 1–27.

3. Reinartz, T. Stages of the discovery process. In Klosgen, W and Zytkow, J (eds) Handbook of Data Mining and Knowledge Discovery. Oxford University Press, 2002. pp. 185–192.

4. Berry, M. J. A., & Linoff, G. S. *Data Mining Techniques : For Marketing, Sales, and Customer Relationship Management* (2nd ed.): John Wiley & Sons. 2004

5. Bellinger G., Casstro D., Mills A. Date, Information, Knowledge, and Wisdom. 2006

6. Losee,R.M. A Discipline Independent Definition of Information. *Journal of the American Society for Information Science*, *48*, 254-269. 1997

7. Ioannidis,Y.E. Query optimization, *ACM Computing Surveys*, vol.28, no.1, 1996, pp.121–123.

8. Chaudhuri, S. An Overview of Query Optimization in Relational Systems. *Proceedings of the ACM PODS*, 1998, pages 34-43.

9. Necib C. B., Freytag J.C.. Ontology based Query Processing in Database Management Systems. Proceedings of the 18th Conference on Advances in Databases and Information Systems (ADBIS'04), Budapest, Hungary, 2004/09 ; 2003, pp. 839-857.

10. Kervalishvili,P., Meparishvili,B. Molecular Machines-Modeling Approaches. ERA-2 Proceedings The Contribution Of Information Technology Science, Economy, Society and Education. T.E.I. of PIREAUS : 453-460 pp. 2008

11. Ediberidze A., Meparishvili B., Janelidze G, New Approaches to a Modeling of Knowledge. IFAC 9 th Workshop on Intelligent Manufacturing Systems (IMS'08), Szczecin, Poland. 99-103 pp. 2008

12. Meparishvili, B. New approach to evolutionary algorithms. .ERA-5 Proceedings The Contribution Of Information Technology Science, Economy, Society and Education. T.E.I. of PIREAUS. 2010.

ონტოლოგიაზე დაფუძნებული მიდგომა სემანტიკური მოთხოვნების ოპტიმიზაციისადმი

ლელა წითაშვილი, ბადრი მეფარიშვილი
საქართველოს ტექნიკური უნივერსიტეტი

რეზიუმე

სტატია ეძღვნება მონაცემთა ბაზების ჰეტეროგენულ მოდელებში მოთხოვნათა დამუშავების მოკლე მიმოხილვასა და აგრეთვე სემანტიკური ძებნისადმი ახალ მიდგომებს. საინფორმაციო და მონაცემთა ბაზების სისტემებში, ინტერნეტში, ვებ-მონაცემთა ბაზებსა თუ სხვა სისტემებში, მონაცემები გროვდება განსხვავებული საინფორმაციო წყაროებიდან ან გენერირდება მომხმარებლების მიერ. მონაცემთა ბაზების მართვის ტრადიციული მიდგომები ჯეროვან ყურადღებას ვერ უთმობენ მომხმარებლის მოთხოვნის სემანტიკის მიმართ. მოთხოვნათა ოპტიმიზაციის მეთოდები ძირითადად იყენებს რელაციური ალგებრის გამოსახულებებში ოპერატორების გადაწყობის მიდგომას მათგან საუკეთესოს შერჩევისა და იმპლემენტაციის თვალსაზრისით. სემანტიკური მოთხოვნების ოპტიმიზაციის კლასის ერთ-ერთ მეთოდს წარმოადგენს ონტოლოგიაზე დაფუძნებული მიდგომა. ჩვენ მიერ წარმოდგენილია მულტიაგენტური სინერგეტიკული მოდელის ევოლუციური ოპტიმიზაციის ახალი მიდგომა.

ОСНОВАННЫЙ НА ОНТОЛОГИИ ПОДХОД ДЛЯ ОПТИМИЗАЦИИ СЕМАНТИЧЕСКИХ ЗАПРОСОВ

Циташвили Л., Мепаришвили Б.
Грузинский Технический Университет

Резюме

Статья посвящена краткому обзору обработки запросов баз данных гетерогенных моделей, а также новым подходам семантического пойска. В информационных системах и базах данных, интернете, Web-базах данных или других системах сбор данных производится из различных источников или генерируется пользователями. Традиционные подходы не обращают должного внимания на семантику запроса пользователя. Методы оптимизации запросов, в основном, используют подход перестраивания операторов реляционной алгебры, выбора из них наилучшего решения и внедрения. Одним из методов класса оптимизации семантических требований является подход, основанный на онтологии. Представлен новый подход эволюционной оптимизации мультиагентной синергетической модели.