Fuzzy Ontology-based Spatial Data Warehouse for Context-aware Search and Recommendation

Hajer Baazaoui-Zghal
Riadi laboratory, University of Manouba, Tunis, Tunisia

Keywords: Fuzzy Ontology, Spatial Data Warehouse, Context-aware Search.

Abstract: The need to build a spatial data warehouse over all structured and unstructured data is becoming necessary in many fields. In addition, contextualized results are considered as key challenges in data warehouses, which are frequently related to only one context. However, in real life applications, decisional data are shared by several users having different profiles, which makes contextual awareness essential for decision support. In this paper, we propose a fuzzy Ontology-based Spatial Data Warehouse for contextual search and recommendation, composed of 3 main layers: (1) Data layer composed of structred, unstructred and users data, (2) Knowledge and context-aware layer and (3) Online application layer. The originality of the work described here consists on integration of uncertain data at different levels of the knowledge layer and having in the same decisional architecture contextual search and recommendation.

1 INTRODUCTION

Decision-makers become more and more exigent face to the exponential growth of storage and large amounts of structured and unstructured data. In this context, data warehouses have been introduced to present solutions for storage, mining and exploration. Indeed, the decision-makers need a knowledge-based warehouse taking into account the context. Thus, integrated approaches based on semantics have been proposed aiming exploration of users knowledge and data sources. Nevertheless, several limits have been observed mainly, related to the lack of contextualization and the integration of uncertain data. A lack of contextual intelligence in case of search in Data Warehouse (DW) is remarked and a need for generation of personalized results is driving the usage of it in context-aware applications. Context consists of all aspects linked to the user and domain that may affect the decision process. Actually data warehouse-based information systems are faced to several challenges, like considering users contexts and preferences, and also considering uncertain data and needs. In fact, vagueness, uncertainty inherently existing in spatial data and imprecision of feature values are not supported by such method. Indeed, fuzzification is integrated into the case-based reasoning process to handle such imprecision and uncertainty.

In this paper, we propose an architecture aiming to assist users during their search for pertinent results and presenting contextual recommendations. Furthermore, our proposal supports integration of imprecise knowledge.

Our motivation is that considering input in the presence of context factors may improve performance and efficiency of the decision process, which is not easily detectable with classic data warehousing methods. Indeed, due to nature of data which is both dynamic and uncertain, data should be interpreted differently depending on current situation (context). Our second motivation relies on the linguistic ambiguity problems, which make crisp ontologies less sufficient when dealing with uncertain knowledge. So, we propose to integrate fuzzy logic into ontologies and contextualization. We bring three main contributions, the proposal (1) integrates different data sources, (2) considers contextual and uncertain data, (3) presents contextual search and recommendations.

The remaining of this paper is organized as follows. Section 2 describes our proposal. Section 3 presents an overview of works related and position our proposal. Finally, Section 3 concludes and proposes directions for future research.

Baazaoui-Zghal, H.
Fuzzy Ontology-based Spatial Data Warehouse for Context-aware Search and Recommendation.
DOI: 10.5220/0006010501610166
In *Proceedings of the 11th International Joint Conference on Software Technologies (ICSOFT 2016) - Volume 2: ICSOFT-PT*, pages 161-166
ISBN: 978-989-758-194-6

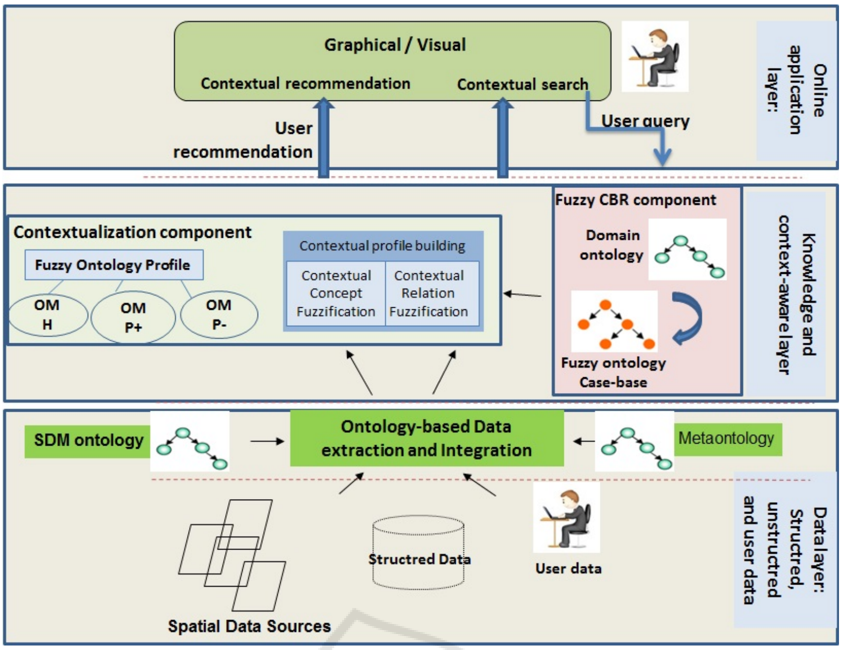

Figure 1: General architecture.

2 FUZZY ONTOLOGY-BASED SPATIAL DATA WAREHOUSE FOR CONTEXT-AWARE SEARCH AND RECOMMENDATION

The objective of the work described in this paper is to propose an architecture that supports both context-aware search and recommendation based on the approach of the data warehouses and data marts which store the relevant data for the decision-makers. We recall that the main steps of data warehouse design starting from heterogeneous sources are: conceptual design, logical design and physical design. The whole process includes the storage, the analysis and the exploration within spatial data, with an aim of improving the process of decision-making. The general architecture of our proposal is given by Figure 1, which shows the workflow for returning answers queries and generating recommendations. It is composed of three main layers: (1) Data layer composed of structred, unstructred and users data, (2) Knowledge and context-aware layer and (3) Online application layer.

2.1 The Data Layer

The data layer allows manipulation of unstructured data, structured data as well as user's data which is

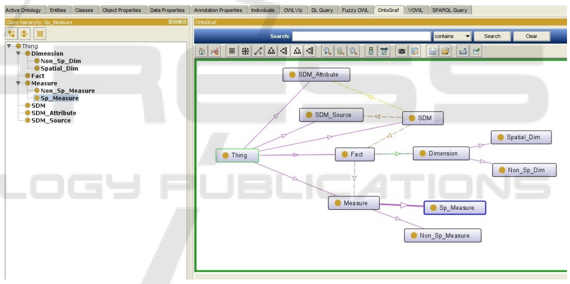

Figure 2: Ontology-based Star Schema.

also considered as input in our proposal as. In a previous work we have proposed a tool for Spatial Data Marts Design and Generation (Zghal et al., 2003), the idea is to create a spatial data warehouse by assembling Spatial Data Marts (SDM). In the current work, from the global sources which can be represented through traditional models dedicated to the management of the spatial data, we propose at first to build the ontology-based Star Schema for SDM building. We recall that ontology is knowledge structure that represents the concepts, relationships, instances and properties for a given domain, which has proven their usefulness to model information systems based on the semantic and knowledge level. In the currect work, we define an ontology for the extraction and integration of data (cf. figure 2). An ontology provides a conceptual representation of the application domain (a shared vocabulary). Then, a selection of attributes from the determined ontological representation and

their correspondence to attributes of the defined star schema in order to populate a specific instance. The meta-ontology (cf. Figure 3) allows modeling and keeping trace about the way the ontology was built; (3) An ontology SDM is characterized by its multidimensional structure formed by facts, dimensions and measurements. Dimensions and measurements can be spatial or nonspatial. We propose to model the hierarchy dependences as a class called Hierarchy. We insist here on the importance to keep a history of the different levels of the hierarchy for dimensions and measurements given the nature of the spatial data.
The fact corresponds to the topic or subject of analysis and is represented by a class. A measurement can be numerical when it contains only numerical data such as the returned monthly one of an area. A measurement can also be spatial: a collection of links to spatial objects. For example the case of a generalization of areas having temperature and precipitations in the same cell. Thus, measurement forms a collection of links to the corresponding areas. So, the global ontology consists of the union of the application ontologies, and a set of axioms that define properties requirements. Our proposal uses ontology as a tool to solve the semantic heterogeneity problem instead of using metadata only.

Four types of ontologies dimensions can be identified: non-geometric spatial ontology dimension, geometric to nongeometric spatial ontology dimension, entirely geometric ontology dimension and temporal ontology dimension.

In general the Spatial dimension (cf. Figure 3) describes the representation of the territory surface, it could comprise specific members, but that would restrict the cartographic representation of the data at one moment given according to only valid territorial cutting to this moment. For the thematic dimensions, several logical models has been be defined. The nongeometric spatial dimension contains nongeometric data, like nominal data making it possible to locate a phenomenon in space. Such a dimension can start with the names of the municipalities and their generalization can, also be non-geometric. The entirely geometric dimension is a dimension of which all the levels and even close generalization are and geometric. Finally, temporel dimension is an unavoidable element in any information systems. Temporal dimension also constitutes strategic information to predict a future behavior or to explain the causes of the current state of the things.

Ontologies allows firstly to define the overall integration schema (global ontology) and the different sources to be integrated. The sources are mapped to a local ontology which will describe the data sources

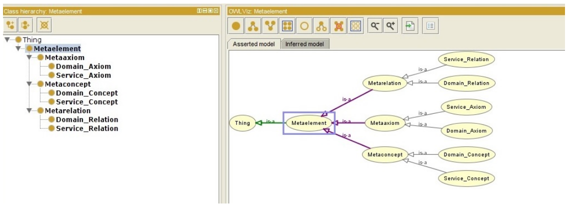

Figure 3: MetaOntology.

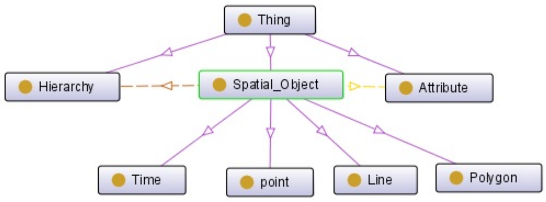

Figure 4: Spatial object description.

semantics. Following are the rules to map a database to ontology:

- The database table is mapped to an ontology class.
- If a database table is related to another, then the two tables are mapped to classes with parents-child relationship. If a database table is related to two tables, then the table is divided into two transferred classes.
- The primary key is mapped to a data type property of the ontology. The foreign key would be mapped to an object property of the ontology.
- The attributes of a table are mapped to properties of the equivalent class.

Our architecture integrates a reusable metaontology which aims to explicitly specify knowledge about the concepts, relationships, instances and axioms extraction, the learned patterns and frames, and the semantic distance, which is generic and could be used in other domains.

The metaontology contributes to:

- Ameliorate the ontology building process by specifying the knowledge that could help the designer to add concepts and relationships,
- Keep trace of the knowledge that led to the insertion of each element in the ontology (concepts, relationships, axioms and instances),
- Maintain the clarity and the coherence of the domain ontology that is dynamically enriched,
- Reuse the knowledge associated to the construction of each ontology (domain, structure and services), It contains knowledge about the representation of each ontology of our architecture (domain, structure and services) and about the axioms.

The output of the *data layer* is an ontology-based Data Warehouse which stores semantics, annotations along with the mechanisms that allow the execution of analysis operations over the stored data, so we obtain a new semi-structured repository with all sources integrated.

2.2 The Knowledge and Context-aware Layer

This layer is mainly composed of a *fuzzy ontology-based Case-Base Reasoning (CBR) component* and *Contextualization component*.

Fuzzy Ontology-based CBR Component: We point out that CBR is a problem-solving method based on the concept of ”case”, which is the description of a problem and its solution. The main idea under CBR consists in storing experiences as cases and problem-solving processes as instances of cases. When a new problem is found, the system uses the relevant past stored cases to interpret or to solve it. The combination of ontologies (as semantic background) and CBR mechanism (to enrich the ontology from search feedback) can improve the performance of Semantic Web search.
In a previous work we proposed to manage and store the cases in a crisp ontology (Elloumi-Chaabene et al., 2011). But crisp ontologies are not able to support uncertain information. One interesting solution is to integrate fuzzy logic into ontology to handle vague and imprecise information. Indeed, this component converts our crisp case base into a fuzzy case-base ontology. We apply the methodology for converting a crisp ontology to a fuzzy ontology proposed (Zghal and Ghézala, 2014). The most critical steps in CBR process are the case representation and case retrieval. We concentrate on these two main steps to improve the performance of the knowledge and contextual process.
A set of all the cases is modelled throw the fuzzy representation. For each fuzzy case, we calculate the weighted fuzzy similarity degree between the current case and all the elements of the fuzzy case set, and select the most similar ones based on maximum similarity degree. Case attributes could be Fuzzy Property Attributes, Fuzzy Valued Property Attributes and Fuzzy Relation Attributes. So, when attributes have uncertain values we need fuzzy ontology-based CBR. The case is retrieved in the fuzzy ontology according to the new problem. the types of records features including numerical, fuzzy, ordinal, lexical, and semantic types. The fuzzy types are represented by a fuzzy ontology, and the semantic types are based on the SDM ontology. The system converts the query case crisp concept into a fuzzy semantic ontology, which passes to the retrieval engine to find the most similar cases. Cases are stored in a fuzzy ontology as concept instances.
During retrieval the fuzzy similarity of a case can is calculated based on a fuzzy membership function for each feature that specifies the desired similarity for any possible difference in the feature values.

Contextualization Component: The context is defined as a set of ontological concepts present in the items recently selected by the user. A contextual fuzzification for the personalized spatial ontology is applied. It begins by extracting the context of the users query from his profile. Then, a contextual fuzzification using Babelnet is performed in order to assign membership values based on the users interests. Context extraction: We consider the context to be a concept extracted from the user profile which is semantically related to the current concept. In order to determine this concept from a given profile, a stop word removal, a lemmatization, POS (Part-Of-Speech) tagging and a matching process with ontology are firstly executed. The output of this process is a set of concepts extracted from the user profile. The most close concept to the concept is considered as the context. In order for the context to be the most representative of the concept, we measure the degree of co-occurrence between these two concepts, used as an estimation of similarity between them. Our component computes the similarity between two concepts in unsupervised manner using the number of results returned by a specialized search engine. The research of similar vocabulary based on text statistics adopts the hypothesis that the context of a word can provide enough information for the word definition. Therefore, the candidate concept which has the higher hit counts with the current concept is considered as the context. The particularity of this extraction method lies in the exploitation of concepts that appear the most in a given corpus which helps extracting relevant concepts the users query and his profile. The contextual fuzzification is mainly based on the fuzzification process of an existing crisp ontology is usually performed using a fuzzification function. The proposed fuzzification function relies on the context ctx in order to favor related concepts to the profile and the concept. The membership value of the relation between a given concept and the current concept, namely contextualized concept.

Fuzzy Ontology Profile: We adopted in this work the ontological structure for the user profile representation. The idea is to define individual ontologies which could be composed later into the same fuzzy ontology.

Definition:
The formal Fuzzy Profile Ontology structure is defined as follows:
$O_{fuz} = \{C,R,A\}$, where C is a set of fuzzy concepts, R is the set of fuzzy relations and A is a set of Axioms expressed in a logical language.
Let us consider an ontology set C={ $O1$, $O2$,... On}, where $O1$,$O2$,...,On are fuzzy profile ontologies.
It is important to note that these fuzzy ontologies are proposed to generate more contextual results.

The proposed fuzzy ontology allows the management of the :

- The historic of user manipulations is represented and the associated dates are also stored in the proposed fuzzy ontology.
- the users' interactions with the system in order to implicitly extract information about his interests and preferences. It is composed of the requested concepts in the made searches and the relations between them. For each concept, the last search's date is also recorded.
  - The positive preferences, which represents the desired concepts and the relations between them, are stored in the same ontology. The user interests are represented by the concepts which are considered relevant in previous search and the relations between them.
  - The negative preferences, which represents the undesired concepts and the relations between them are also stored in the same ontology. Storing the concepts which are considered irrelevant in previous search and the relations between them, allow the IR system to avoid returning results containing these concepts.

Contextual User Profile Building: The contextual user profile building step is designed to extract concepts and relations and assign to them membership values in order to enrich the proposed profile ontology with fuzzy concepts and relations. A knowledge extraction step is firstly done by analyzing results in order to extract concepts and relations (both taxonomic and semantic) from resulted documents (both relevant and irrelevant document groups). Then, a contextual fuzzification process is applied in order to assign membership values to each concept and relation. Our contextualized concept fuzzification method favors the most appearing concepts. It also favors the concepts that appear with the right context since it relies on the sum of weights of the concept and its context. Therefore the memberships values of contextualized concepts are higher than the ones without their context.

Recommendation Component: this component uses the records of previous similar experiences to generate suggested or create new items when no existing ones meet the needs or preferences of the user. In a previous work a comparative study (Haddad et al., 2015) concerning the proposed approaches of recommendation has shown that Content-Based Filtering (CBF) have the better results when considering a quality measure. In our architecture the CBF engine is mainly used to adapt the recommendations to the preferences of the users and ensure a degree of diversity and novelty in the suggested recommendations. The user preferences are represented as vectors, the intensity of the interest of user for a given concept (a class or an instance) in a the ontology is measured taking into account its positive or negative status described in the previous section.

2.3 The Online Application Layer

The application layer provides results of the previous components to the user during decision support process. The search process begins when the user submits a query or when recommendation is generated by the knowledge layer. In order to show that our proposal can have a great interest for contextual search and recommendation, the architecture has been implemented. From more technical point of view the frameworks that have been employed in the implementation of the fuzzy ontology-based CBR are mainly PostgreSQL, PostGIS and Java-based prototype using Fuzzy Owl (Bobillo et al., 2013)(for managing fuzzy ontologies).

3 RELATED WORKS AND DISCUSSION

Different proposals have been made regarding how to represent a conceptual multidimensional schema to model data warehouses and data marts. Different categorization of data models and comparisons of several multidimensional data models are proposed in the literature, where the main identified levels were: conceptual, logical, physical and formal.
On the one hand, extensions were made to make appropriate the analysis and the algorithms to specificities of the handled spatial data . On the other hand, several context models have been proposed to support

users context (Khouri et al., 2013),...
Recent works proposed to describe data sources by using global and local ontologies, for more semantic data warehouses (Bellatreche et al., 2013) (Cuzzocrea and Simitsis, 2012). Indeed, several limits have been observed, mainly related to the lack of integration of: uncertain data for decision support, users preferences, and their contexts and preferences. Crisp ontologies are not capable to support uncertain information and integration of fuzzy logic into ontology to handle vague and imprecise information has proven its usefulness. Moreover, Case Based Reasoning is an important field which has been applied to various problems (Elloumi-Chaabene et al., 2011). To the best of our knowledge among existing works, this is the first time that context, CBR and ontologies are integrated together in a decision support system which consider uncertain data at different levels (data sources, semantics and cases)

4 CONCLUSION AND PERSPECTIVES

In this paper, we proposed a Fuzzy Ontology-based Spatial Data Warehouse for contextual search and recommendation. Our proposal takes place through three main layers: Data layer, Knowledge and context-aware layer and Online application layer. Unlike many previous decision support approaches, the originality of this work is that it takes into account imprecise data at two different levels: fuzzy profile ontology and fuzzy ontology-based CBR. The architecture has been implemented and an evaluation of the retrieval tasks is currently conducted.

REFERENCES

Bellatreche, L., Khouri, S., and Berkani, N. (2013). Semantic data warehouse design: From ETL to deployment à la carte. In *Database Systems for Advanced Applications, 18th International Conference, DASFAA 2013, Wuhan, China, April 22-25, 2013. Proceedings, Part II*, pages 64–83.

Bobillo, F., Delgado, M., and Gómez-Romero, J. (2013). Reasoning in fuzzy OWL 2 with delorean. In *Uncertainty Reasoning for the Semantic Web II, International Workshops URSW 2008-2010 Held at ISWC and UniDL 2010 Held at FLoC, Revised Selected Papers*, pages 119–138.

Cuzzocrea, A. and Simitsis, A. (2012). Searching semantic data warehouses: models, issues, architectures. In *Proceedings of the 2nd International Workshop on Semantic Search over the Web, Istanbul, Turkey, August 27, 2012*, page 6.

Elloumi-Chaabene, M., Mustapha, N. B., Zghal, H. B., Moreno, A., and Sánchez, D. (2011). Semantic-based composition of modular ontologies applied to web query reformulation. In *ICSOFT 2011 - Proceedings of the 6th International Conference on Software and Data Technologies, Volume 1, Seville, Spain, 18-21 July, 2011*, pages 305–308.

Haddad, M. R., Zghal, H. B., Ziou, D., and Ghézala, H. H. B. (2015). Towards a generic architecture for recommenders benchmarking. In *ICAART 2015 - Proceedings of the International Conference on Agents and Artificial Intelligence, Volume 2, Lisbon, Portugal, 10-12 January, 2015.*, pages 435–442.

Khouri, S., Saraj, L. E., Bellatreche, L., Espinasse, B., Berkani, N., Rodier, S., and Libourel, T. (2013). Cidhouse: Contextual semantic data warehouses. In *Database and Expert Systems Applications - 24th International Conference, DEXA 2013, Prague, Czech Republic, August 26-29, 2013. Proceedings, Part II*, pages 458–465.

Zghal, H. B., Faïz, S., and Ghézala, H. H. B. (2003). CASME: A CASE tool for spatial data marts design and generation. In *Design and Management of Data Warehouses 2003, Proceedings of the 5th Intl. Workshop DMDW'2003, Berlin, Germany, September 8, 2003*.

Zghal, H. B. and Ghézala, H. B (2014). A fuzzy-ontology-driven method for a personalized query reformulation. In *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2014, Beijing, China, July 6-11, 2014*, pages 1640–1647.