République Tunisienne
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

---

DOSSIER DE CANDIDATURE

Soumis à

La commission Nationale de Recrutement des Maîtres assistants

Dans la discipline

Informatique

Publications Scientifiques

Elaboré par :

Dr. Ghassen HAMDI

Laboratoire MARS, Université de Sousse

I. Table des matières

I. Articles publiés dans des Journaux internationaux 4

1. Journal of King Saud University-Computer and Information Sciences 4

2. Annals of Mathematics and Artificial Intelligence 9

3. Scientific Journal of King Faisal University Basic and Applied Sciences 13

II. Articles soumis dans des Journaux Internationaux 17

1. Social Network Analysis and Mining 17

2. ACM Transactions on Database Systems 21

3. Computational and Applied Mathematics 25

4. Computer Science and Information Systems 29

III. Articles publiés dans des Conférences Internationales 33

1. International Conference on Cyberworlds (CW 2023) 33

2. International Conference on Innovations in Intelligent Systems and Applications (INISTA 2023) 37

3. International Conference on Database and Expert Systems Applications (DEXA 2021) 41

4. International Conference on New Trends in Intelligent Software Methodologies, Tools and Techniques (SOMET 2021) 44

5. International Symposium on Artificial Intelligence and Mathematics (ISAIM 2018) 48

Liste des publications Scientifiques

Articles publiés dans des Journaux internationaux

[J1] **Hamdi Ghassen,** Telli Abdelmoutia and Omri Mohamed Nazih (2022). "Querying of several DL-Lite knowledge bases from various information sources-based polynomial response unification approach". Journal of King Saud University-Computer and Information Sciences, **Vol. 34, No.4, pp.271–299**. ISSN: 2213-1248 **(IF=6.9, Quartile=Q1), doi:** https://doi.org/10.1016/j.jksuci.2020.06.002**.**

[J2] **Hamdi Ghassen,** Omri Mohamed Nazih, Benferhat Salem, Bouraoui Zied and Papini Odile (2021). "Query answering DL-lite knowledge bases from hidden datasets". *Annals of Mathematics and Artificial Intelligence*, **Vol. 89, No.3-4, pp. 271--299.** ISSN: 1573-7470

(IF=1.2, Quartile=Q3), doi: https://doi.org/10.1007/s10472-020-09714-2.

[J3] Telli Abdelmoutia**, Hamdi Ghassen** and Omri Mohamed Nazih (2021). "Lexicographic repair under querying prioritised dl-lite knowledge bases". *Scientific Journal of King Faisal University Basic and Applied Sciences*, **Vol. 22, No.1.** ISSN: 16580311

(IF=0.149, Quartile=Q3), doi: 10.37575/B/SCI/0054**.**

Articles soumis dans des Journaux Internationaux

[J1] **Hamdi Ghassen** and Omri Mohamed Nazih (2023). "Personalization of OLAP queries for hierarchical visualization under constraints". *Social Network Analysis and Mining*, **(Under Review). ISSN: 1869-5469 (IF=2.8, Quartile=Q1)**

[J2] **Hamdi Ghassen,** Telli Abdelmoutia and Omri Mohamed Nazih (2023)."A Mechanism for Generating OLAP Queries Dynamically based on User Preferences and Hierarchical Visualization Constraint", ACM Transactions on Database Systems, **(Under Review)** ISSN: 0362-5915 **(IF=1.09, Quartile=Q1)**

[J3] **Hamdi Ghassen**, Telli Abdelmoutia**,** and Omri Mohamed Nazih (2024)."Enhancing DL-Lite Possibilistic Logic: Interval-Based Knowledge Representation and Reasoning with Degrees of Inconsistency", Computational and Applied Mathematics, **(Under Review)** ISSN: 1807-0302 **(IF=2.6, Quartile=Q2)**

[J4] **Hamdi Ghassen,** (2024)."Incremental and Hierarchical Visualization of Query Results on a Data Warehouse", *Computer Science and Information Systems*, **(Under Review)** ISSN: 2406-1018 **(IF=1.17, Quartile=Q3)**

Articles Publiés dans des Conférences Internationales

[C1] **Hamdi Ghassen** and Omri Mohamed Nazih., "A Personalized Multidimensional Navigation in a Limited Visualization Context ". International Conference on Cyberworlds (**CW** 2023), 3-5 October 2023. Sousse, Tunisia, pp. 54--61.

doi : 10.1109/CW58918.2023.00018, (**Classe C selon Core 2023)**

[C2] **Hamdi Ghassen** and Omri Mohamed Nazih, "Personalized OLAP queries under Hierarchical Visualization Constraint". International Conference on Innovations in Intelligent Systems and Applications (**INISTA**'23). 20-23 September 2023. Hammamet, Tunisia, pp. 1-6.

doi: 10.1109/INISTA59065.2023.10310545, (**Classe C selon Core 2023**)

[C3] **Hamdi Ghassen** and Omri Mohamed Nazih, "Towards New Model for Handling Inconsistency Issues in DL-Lite Knowledge Bases". The 32nd International Conference on Database and Expert Systems Applications (**DEXA** 2021), 27-30 September 2021. Virtual Event, pp. 93-99.

doi : 10.1007/978-3-030-86475-0_10, (**Classe B selon Core 2021**)

[C4] **Hamdi Ghassen** and Omri Mohamed Nazih, "New Model for Handling Inconsistency Issues in DL-Lite Knowledge Bases". The 20th International Conference on New Trends in Intelligent Software Methodologies, Tools and Techniques (**SOMET** 2021), 21-23 September 2021. Cancun, Mexico, pp. 183--196.

doi : 10.3233/FAIA210019**,** (**Classe C selon Core 2021**)

[C5] **Hamdi Ghassen,** Omri Mohamed Nazih, Papini Odile, Benferhat Salem and Zied Bouraoui "Querying DL-lite Knowledge Bases from Hidden Datasets", International Symposium on Artificial Intelligence and Mathematics (**ISAIM** 2018), January 3-5, 2018. Fort Lauderdale, Florida, USA. (**Classe B selon Core 2018)**

I. Articles publiés dans des Journaux internationaux

1. Journal of King Saud University-Computer and Information Sciences

Remarque : Cet article est accepté **1 juin 2020** et publié en ligne **12 juin 2020**. Le journal est devenu payant après **31 octobre 2021.** Voir la page d'accueil de la revue (Journal of King Saud University - Computer and Information Sciences | ScienceDirect.com by Elsevier)

Article [J1] : Querying of several DL-Lite knowledge bases from various information sources-based polynomial response unification approach.

Auteurs : **Hamdi Ghassen,** Telli Abdelmoutia and Omri Mohamed Nazih

Journal : Journal of King Saud University-Computer and Information Sciences

Etat : Publié

Quartile : Q1

SJR : 0.924

ISSN: 2213-1248

ScienceDirect

Journal of King Saud University - Computer and Information Sciences

Open access

6.9 CiteScore

6.9 Impact Factor

Submit your article | Guide for authors

Menu | Search in this journal

About the journal

Peer Review under the responsibility of King Saud University

Announcement: In **2022** the *Journal of King Saud University - Computer and Information Sciences* will become an author paid open access journal. Authors who submit their manuscript after **October 31st 2021** will be asked to pay an Article ...

ScienceDirect

Journal of King Saud University - Computer and Information Sciences

Volume 34, Issue 4, April 2022, Pages 1161-1171

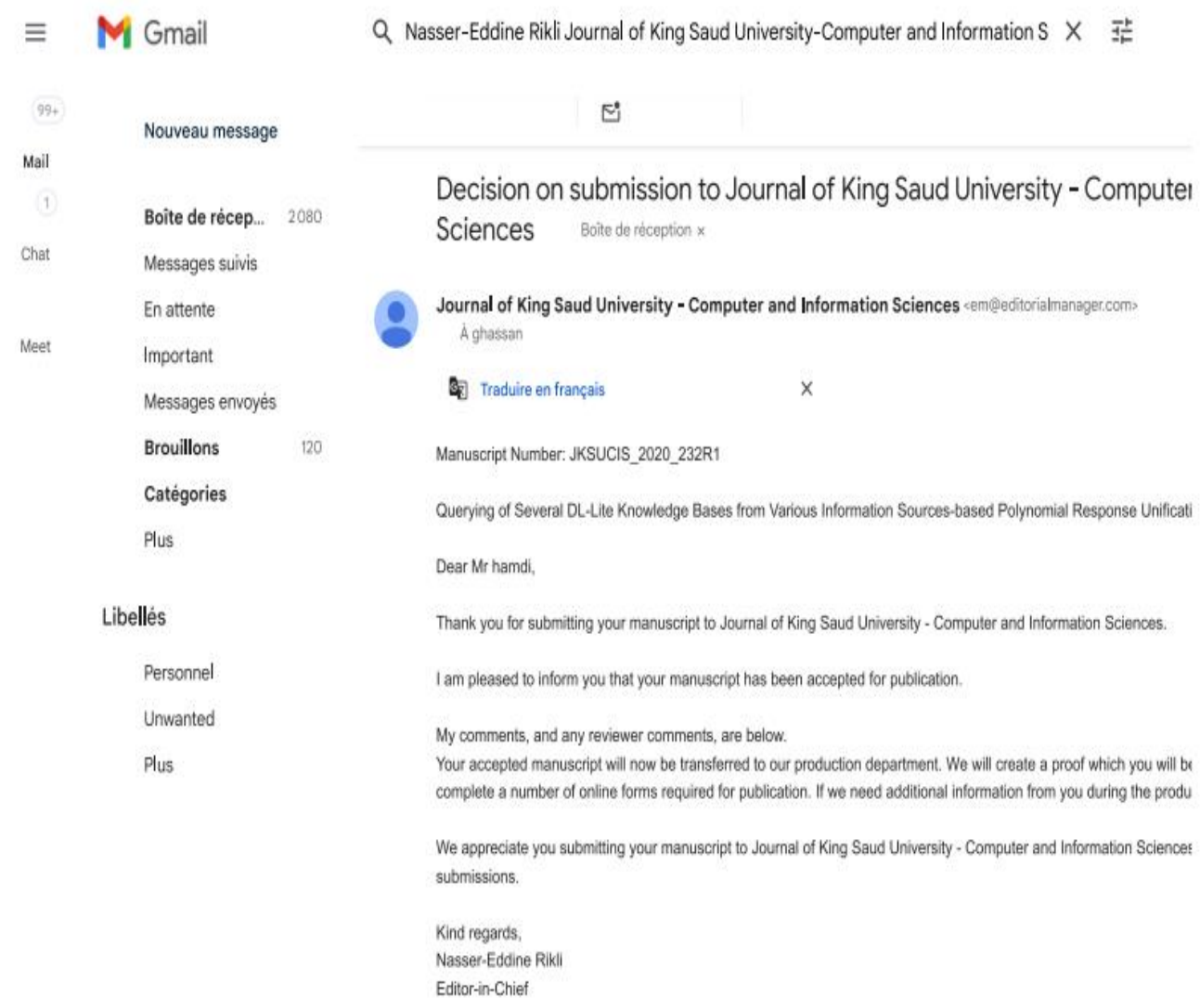
Querying of several DL-Lite knowledge bases from various information sources-based polynomial response unification approach

Ghassen Hamdi [a], Abdelmoutia Telli [b], Mohamed Nazih Omri [a]

[a] MARS Research Laboratory, ISITCOM, University of Sousse, Tunisia

[b] Biskra University, Algeria

Received 14 February 2020, Revised 27 May 2020, Accepted 1 June 2020, Available online 12 June 2020, Version of Record 29 March 2022.



Gmail

Nasser-Eddine Rikli Journal of King Saud University-Computer and Information S

99+

Mail

1

Chat

Meet

Nouveau message

Boîte de récep... 2080

Messages suivis

En attente

Important

Messages envoyés

Brouillons 120

Catégories

Plus

Libellés

Personnel

Unwanted

Plus

Decision on submission to Journal of King Saud University - Computer Sciences Boîte de réception ×

Journal of King Saud University - Computer and Information Sciences <em@editorialmanager.com>
À ghassan

Traduire en français

Manuscript Number: JKSUCIS_2020_232R1

Querying of Several DL-Lite Knowledge Bases from Various Information Sources-based Polynomial Response Unificati

Dear Mr hamdi,

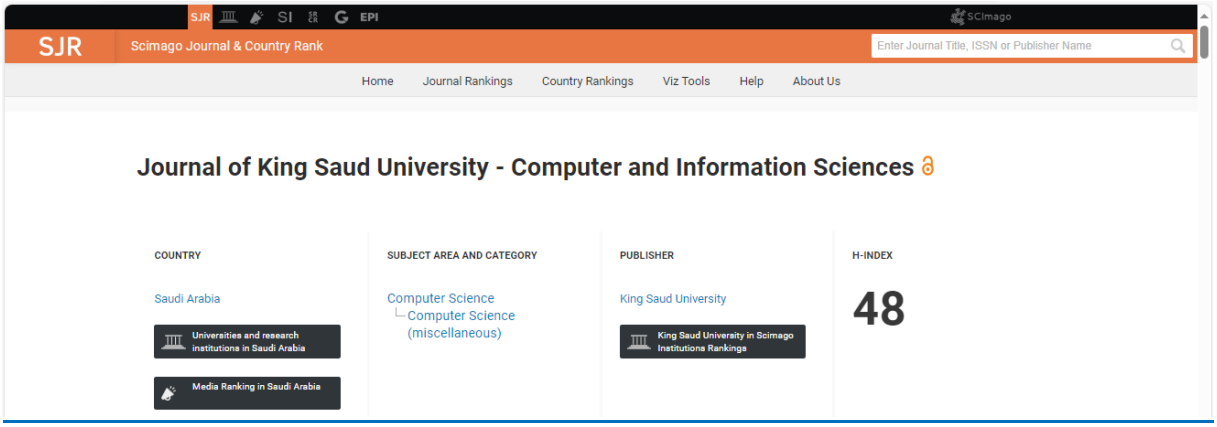
Thank you for submitting your manuscript to Journal of King Saud University - Computer and Information Sciences.

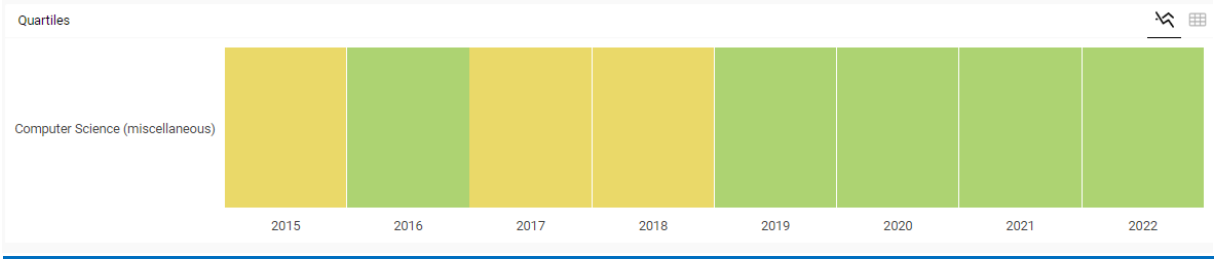
I am pleased to inform you that your manuscript has been accepted for publication.

My comments, and any reviewer comments, are below.
Your accepted manuscript will now be transferred to our production department. We will create a proof which you will be complete a number of online forms required for publication. If we need additional information from you during the produ

We appreciate you submitting your manuscript to Journal of King Saud University - Computer and Information Sciences submissions.

Kind regards,
Nasser-Eddine Rikli
Editor-in-Chief


SJR
Scimago Journal & Country Rank
Enter Journal Title, ISSN or Publisher Name
Home
Journal Rankings
Country Rankings
Viz Tools
Help
About Us
Journal of King Saud University - Computer and Information Sciences
COUNTRY
Saudi Arabia
Universities and research institutions in Saudi Arabia
Media Ranking in Saudi Arabia
SUBJECT AREA AND CATEGORY
Computer Science
Computer Science (miscellaneous)
PUBLISHER
King Saud University
King Saud University in Scimago Institutions Rankings
H-INDEX
48


Quartiles
Computer Science (miscellaneous)
2015
2016
2017
2018
2019
2020
2021
2022

Journal of King Saud University - Computer and...
Q1
Computer Science (miscellaneous)
best quartile
SJR 2022
0.92
powered by scimagojr.com

Journal of King Saud University – Computer and Information Sciences 34 (2022) 1161–1171

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Querying of several DL-Lite knowledge bases from various information sources-based polynomial response unification approach

Ghassen Hamdi [a], Abdelmoutia Telli [b], Mohamed Nazih Omri [a]

[a] MARS Research Laboratory, ISITCOM, University of Sousse, Tunisia
[b] Biskra University, Algeria

ARTICLE INFO

Article history:
Received 14 February 2020
Revised 27 May 2020
Accepted 1 June 2020
Available online 12 June 2020

Keywords:
Ontology based data access
DL-Lite
Query answering
Inconsistent answers
Repairs answers
Prioritized knowledge bases

ABSTRACT

The inconsistency in the hierarchical knowledge base is due to the fact that the assertions (ABoxes) come from several sources with different levels of reliability. We introduce the management of this inconsistency problem to query the inconsistent DL-Lite knowledge bases. Most of the approaches studied in the literature first repair all of the inconsistent assertions in the DL-Lite incoherent knowledge base and then question it. In this article, we proceed directly to a knowledge base query to retrieve an exhaustive list of responses and then, once these responses are retrieved, we repair them. The main contribution of this work therefore consists of studying several response strategies to queries from various data sources using a recurrent function to calculate the rank of coherence which will allow us to manage inconsistencies in the set of responses. We study these strategies from different points of view: productivity, precision, recall, F-measure, Accuracy, G-mean and computational complexity. The experiments carried out, on different corpora, and the analysis of the results, which we carried out, show encouraging results in productivity since our approach is much more productive, providing the greatest number of responses, while remaining the best in time compared to the other studied approaches.
© 2020 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Description Logics (DLs) is a recent family of knowledge representation (KR) formalisms to represent the knowledge of an application domain by first defining the relevant concepts of the domain (its terminology), then using these to specify the properties of objects and individuals present in the domain. One of the characteristics of this type of logic is that, unlike some of their predecessors, they have a formal semantics based on logic. Another feature is the emphasis on reasoning as a central service for inferring knowledge implicitly from knowledge that is explicitly contained in the knowledge base (KB). DLs support the models of inference that occur in many applications of intelligent systems, also used by humans for the classification of concepts and individuals.

The classification of concepts determines the subconcept/superconcept relationships, called subsumption relationships in DL, between the concepts of a given terminology, and thus makes it possible to structure the latter in the form of a subsumption hierarchy. This hierarchy provides useful information about the connection between different concepts and can be used to speed up other inference services. The classification of individuals, or objects, determines whether a given individual is always an instance of a certain concept, in other words if this instance relationship is implied by the description of the individual and the definition of the concept. It thus provides useful information about the properties of an individual. Add to that, instance relationships can trigger the application of rules that insert additional facts into the knowledge base.

Description Logics are also formal frameworks for representing and reasoning with ontologies. The DL knowledge base consists of: TBox as a terminological base represents the conceptual knowledge of a particular domain and ABox as an assertional base contains facts or assertions concerning particular individuals Baader et al., 2010.

Recently, Ontology Based Data Access (OBDA) (Maurizio et al., 2011; Poggi et al., 2008; Rodriguez-Muro et al., 2013) is collecting great attention as a new paradigm in which structured knowledge or the ontological view (*i.e.*, stored in a TBox) is used to provide

E-mail addresses: a.telli@univ-biskra.dz (A. Telli), Mohamednazih.omri@eniso.u-sousse.tn (M.N. Omri)
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2020.06.002
1319-1578/© 2020 The Author(s). Published by Elsevier B.V. on behalf of King Saud University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

better exploitation of assertions (*i.e.,* stored as an ABox) when querying them. A crucially important service provided by an OBDA system is query answering, which aims to calculate the answers of a query presented in terms of ontologies.

In many applications, assertions are provided by several potentially conflicting sources with different levels of reliability. In addition, a given source may have different sets of uncertain assertions that together form a prioritized or a stratified assertional base (i.e ABox). A major problem arises in the context of *DL-Lite* query answering is how to deal with the case of inconsistency between ABox and TBox. Indeed, the TBox is generally verified and validated as long as the ABox will typically be larger and may have several modifications and thus may be in contradiction with the TBox.

Many works in the context of OBDA, inspired by database approaches (*e.g.* Arenas et al., 1999; Bertossi, 2011; Chomicki, 2006) or propositional logic approaches (*e.g.* Benferhat et al., 1993; Benferhat et al., 1997; Nebel, 1994), deal with the problem of querying inconsistent DL KB s by proposing several inconsistency-tolerant inferences, called semantics, and were introduced for the lightweight description logic *DL-Lite* (*e.g.* Lembo et al., 2015). Among these semantics, one can quote the ABox Repair (AR) and Intersection ABox Repair (IAR) semantics (Lembo et al., 2010) which are the most known and the most studied. These two semantics are based on the notion of maximal assertional repair, which is based on the notion of repair in the database domain or the maximal consistent subsets in the propositional logic setting (Rescher and Manor, 1970; Brewka, 1989).

Motivation and contribution:

The work we propose in this article poses the problem of inconsistency, in the priority DL-LITE knowledge bases, compared to the rapid and diversified evolution of the amounts of information in this type of knowledge base. It also aims to give an overview of the main work on the problematic and to present the main research contributions made. These motivations and contributions are summarized in the following points:

- Present a summary of the problem of inconsistency in DL-LITE knowledge bases,
- Conduct a review of the literature on inconsistency in the DL-LITE knowledge bases to better understand the principle of the main approaches studied in the literature on this issue,
- Identify the main characteristics of this type of knowledge base, which have not been addressed in the scientific literature by existing approaches,
- Determine the limits of current approaches and their causes,
- Develop a fairly generic strategy that reflects the different aspects related to the problem of inconsistency in DL-LITE knowledge bases,
- Propose a new approach to querying several DL-LITE knowledge bases from various sources of information.

The rest of this article is organized as follows: next Section 2 presents the related works. Section 3 provides a brief refresher on *DL-Lite* and querying multiple prioritized sources. Section 4 explains the notion of inconsistency tolerant reasoning for assertions associated with the answers profile (Conflicts set repair answers, Free set answers, reparis answers, and consistency rank). Section 5 presents the main purpose of this work and gives justifications for each choice. Section 6 provides a computational analysis. In the final section, we summarize the work and we present some future works.

2. Related works

The principal inspiration for the present paper comes from a line of research in inconsistency-handling. Inconsistency is defined with respect to some assertions that contradict the terminology. Typically, a TBox is usually verified and validated while the assertions can be provided in large quantities by various and unreliable sources and may contradict the TBox. Moreover, it is often too expensive to manually check and validate all the assertions. This is why it is very important in OBDA (Ontology-based Data Access) to reason in the presence of inconsistency. Many works (*e.g.* Lembo et al., 2010; Bienvenu and Rosati, 2013), basically inspired by database approaches (*e.g.* Bertossi, 2011), tried to deal with inconsistency in *DL-Lite* by adapting several inconsistency-tolerant inference methods. In many applications, assertions are often provided by several and potentially conflicting sources having different reliability levels.

Moreover, a given source may provide different sets of uncertain assertions with different confidence levels. Gathering such sets of assertions gives a prioritized or a stratified assertional base. The role of priorities in handling inconsistency is very important and it is largely studied in the literature within propositional logic setting (*e.g.* Baral et al., 1991; Benferhat et al., 1995). Several works (*e.g.* Staworko et al., 2012; Du et al., 2013) studied the notion of priority when querying inconsistent databases or DL KBs. Unfortunately, in the OBDA setting, there are only few works, such as the one given in Bienvenu et al. (2014) for dealing with reasoning under prioritized *DL-Lite* ABox.

A recent line of work studies the inconsistency in lightweight ontologies. For instance, the authors in Lembo et al. (2010, 2015), Bienvenu and Rosati (2013), Benferhat et al. (2016) and Baget et al. (2016) investigate the problem of inconsistency in KB s by computing a set of consistent subsets of assertions called repairs, which recovers the consistency with respect to the ontology, and then using them to answer the queries. Moreover, the authors propose in Benferhat et al. (2015), Telli et al. (2017) polynomials algorithms for select a single preferred repair from a prioritized inconsistent *DL-Lite* KB to allow an efficient query answering once the repair is computed. Particularly, the authors in Benferhat et al. (2015) propose a new approach based on the selection of only one preferred repair. However, in Telli et al. (2017), the authors propose a sequential inference strategies based on the selection of one consistent assertional base. The authors in Boughammoura and Omri (2017), Boughammoura et al. (2012) and Boughammoura et al. (2015) propound a new algorithm which makes easy to query answering without access to Web databases.

In addition, the authers in Bienvenu et al. (2016) investigated the computational complexity of reasoning about optimal repair plans (a repair plan is deletions and additions to find a set of ABox modifications) and they proposed interactive algorithms for computing such plans. The authers in Bienvenu et al. (2019) defined explanations for positive and negative answers under the brave, AR and IAR semantics and studied the computational properties of explanations in DL-Lit$_R$. Also, the authers in Trivela et al. (2019) proposed a general framework which considers the ICAR (Intersection Closed ABox Repair) semantics and is based on a rewriting algorithm that can be applied over arbitrary DLs for answering queries over inconsistent DL knowledge bases. However, the authors in Dixit (2019) proposed CAvSAT (Consistent Answering via Satisfiability) as a novel approach that has a potential to build a comprehensive and scalable consistent query answering system.

In our work, a recursive algorithm starts by querying all the whole knowledge base which will allow us to have an exhaustive list of all the possible answers. Then, if this list of answers is inconsistent, the algorithm repairs it. Hence, it do not any correction except if the set answers is inconsistent. However, existing algorithms start by repartion the knowledge base, then, querying this reparation with a sequential function.

3. Syntax and semantics of prioritized *DL-Lite* knowledge base

Let N_C, N_R and N_I, three pairwise disjoint sets of atomic concepts, atomic roles and individuals respectively. Let $A \in N_C, P \in N_R$. Let also '$\neq g$', '$\exists$' and '$^-$' three connectors are used to define complex concepts and roles. *DL-Lite* concepts are defined as follows:

$$B \longrightarrow A \mid \exists R \qquad C \longrightarrow B \mid \neq gB$$
$$R \longrightarrow P \mid P^- \qquad E \longrightarrow R \mid \neq gR$$

where A is an atomic concept, P is an atomic role and P^- is the inverse of an atomic role, B (*resp.* C) is the set of basic (*resp.* complex) concepts, and R (*resp.* E) is the set of basic (*resp.* complex) roles.

Let $\mathcal{K}$=$\mathcal{T}, \mathcal{A}$ be a *DL-Lite* KB where $\mathcal{T}$ is called the TBox and $\mathcal{A}$ is called the ABox. A TBox includes a finite set of inclusion axioms on concepts and on roles respectively of the form inclusion assertions: $B \sqsubseteq C$ (*resp.* negative inclusion assertions $B \sqsubseteq \neq gC$) means that concept B is included in concept C (*resp.* concept B is not included in the concept C) and $R \sqsubseteq E$ (*resp.* $R \sqsubseteq \neq gE$) means that role R is included in the role E (*resp.* role R is not included in role E).

The ABox contains a finite set of assertions (facts) of the form $A(a)$ and $P(a,b)$ where $A \in N_C, P \in N_R$ and $a, b \in N_I$. The semantics is given in terms of interpretations $\mathcal{I}$=$(\Delta^{\mathcal{I}}, .^{\mathcal{I}})$ which consist of an non-empty domain $\Delta^{\mathcal{I}}$ and an interpretation function $.^{\mathcal{I}}$ that assigns to each $a \in N_I$ an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, to each $A \in N_C$ a subset $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and to each $P \in N_R$ an $P^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. A TBox $\mathcal{T}$ is said to be incoherent if there exists a concept C s.t $\forall \mathcal{I} : \mathcal{I} \models \mathcal{T}$, we have $C^{\mathcal{I}}$=$\varnothing$. A *DL-Lite* KB $\mathcal{K}$ is said to be inconsistent if it does not admit any model.

The prioritized profile is a multiset of prioritized or stratified ABoxes denoted by $\mathcal{K}_P = \mathcal{T}, P_s$ where $\mathcal{T}$ is a standard *DL-Lite* TBox and $P_s = \{L_1, ..., L_m\}$ is a prioritized ABox profile. It is assumed that each ABox $L_i \in P_s$ is consistent with the ontology (TBox). In this case, each sets L_i are called layers or strata, which each layer L_i contains a set of assertions with the same level of priority i and they are considered more reliable than those present in a layer L_j when $j > i$. Accordingly, L_1 contains the most important assertions as long as L_m contains the least important ones.

A query is a first-order logic formula, denoted q=$\{(x) \mid \phi(x)\}$, where (x)=$(x_1,\ldots,x_n)$ are free variables, n is the arity of q and atoms of $\phi(x)$ are of the form $A(t_i)$ or $P(t_i, t_j)$ with $A \in N_C$ and $P \in N_R$ and t_i, t_j are terms, *i.e.,* constants of N_I or variables. When $\phi(x)$ is of the form $\exists(y).conj(x,y)$ where y are bound variables called existentially quantified variables, and $conj(x,y)$ is a conjunction of atoms of the form $A(t_i)$ or $P(t_i, t_j)$ with $A \in N_C$ and $P \in N_R$ and t_i, t_j are terms, then q is said to be a conjunctive query (CQ). An answer to a CQ $q(x) \leftarrow conj(x,y)$ over a KB $\mathcal{K} = \mathcal{T}, \mathcal{A}$ is a non empty set of tuples $s = (s_1, \cdots, s_k) \in N^{\mathcal{I}} \times \cdots \times N^{\mathcal{I}}$ such that $\mathcal{T}, \mathcal{A} \models q(s)$.

Let $q(x)$ be a query, we consider $S_{P_s} = \{S_1, \ldots, S_m\}$ a various sets of answers to a query $q(x)$ regarding the prioritized profile P_s called the profile of sets of answers where each S_i is the set of answers to $q(x)$ w.r.t. L_i for $1 \leqslant i \leqslant m$, defined as follows: $S_i = \{s \in N^{\mathcal{I}} \times \cdots \times N^{\mathcal{I}} : \mathcal{T}, L_i \models q(s)\}$. Certainty, when there is no answer to the query $q(x)$ with respect to L_i, S_i = $\varnothing$.

4. Inconsistency tolerant reasoning

In this section, we have assumed that each ABox is consistent with the TBox. Coping with inconsistency can be done by first computing the set of consistent subsets of assertions associated to a set of answers to a given query, called repairs answers.

Let $\mathcal{K}_P = \mathcal{T}, P_s$ be a prioritized *DL-Lite* KB with: $P_s = \{L_1, \ldots, L_m\}, q(x)$ be a query, $S_{P_s} = \{S_1, \ldots, S_m\}$ a set of answers to a query $q(x)$ with respect to $P_s, q_{P_s} = (q_{L_1}, \ldots, q_{L_m})$ is the set of assertions associated to S_{P_s} such that $q_{L_i} = \{q(s) : s \in S_i\}$ where q_{L_i} a set of answers to the query $q(\vec{x})$ with respect to each L_i. (See Fig. 2).

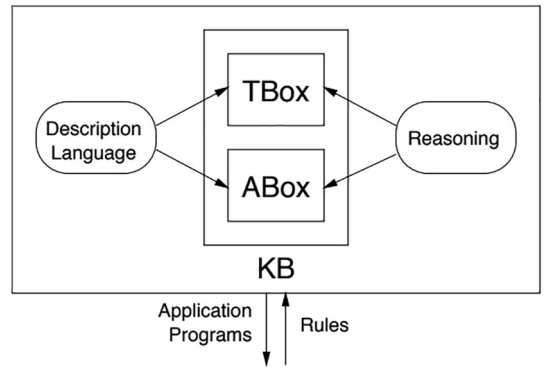


Fig. 1. Architecture of a knowledge representation system based on description logics.

4.1. Conflicts set answers

The conflicts sets answers represent a minimal inconsistent subset $\mathcal{C}$ of the assertions associated to S_{P_s} such that $\langle \mathcal{T}, \mathcal{C} \rangle$ is inconsistent. (See Fig. 3),

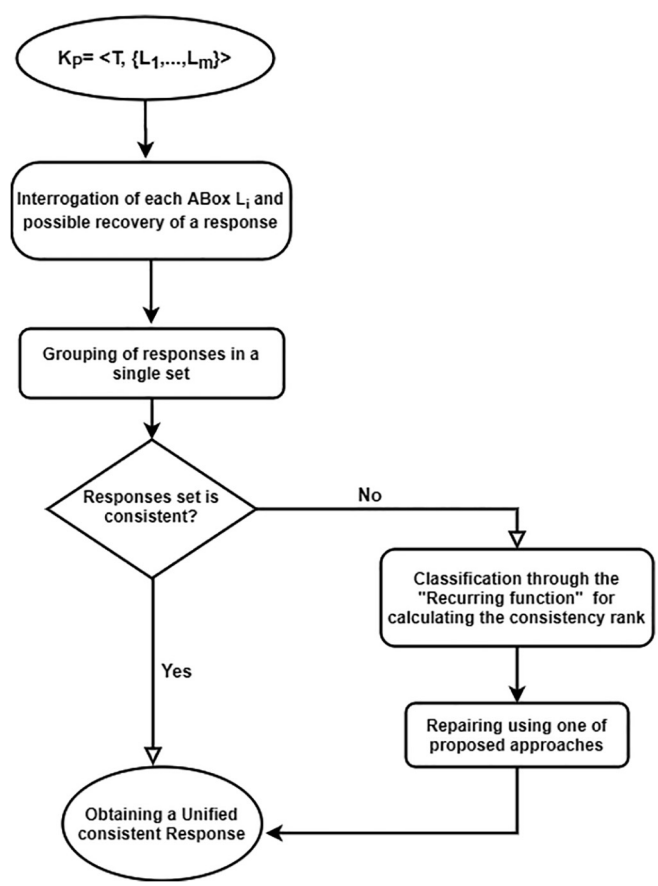


Fig. 2. Response processing process from various inconsistent DL-Lite knowledge bases.

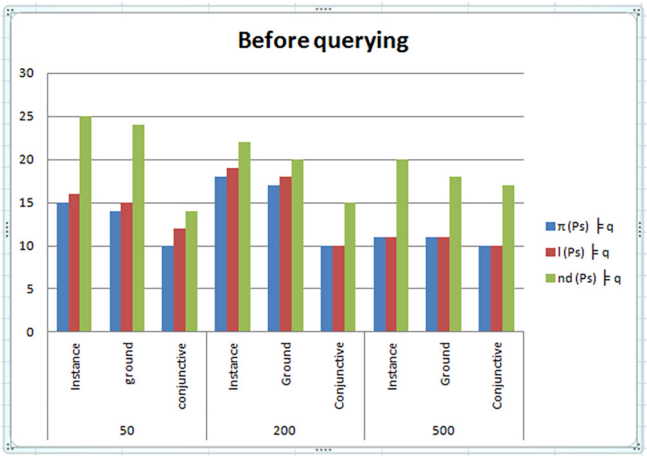


Fig. 3. Productivity of applying our repair algorithms before querying.

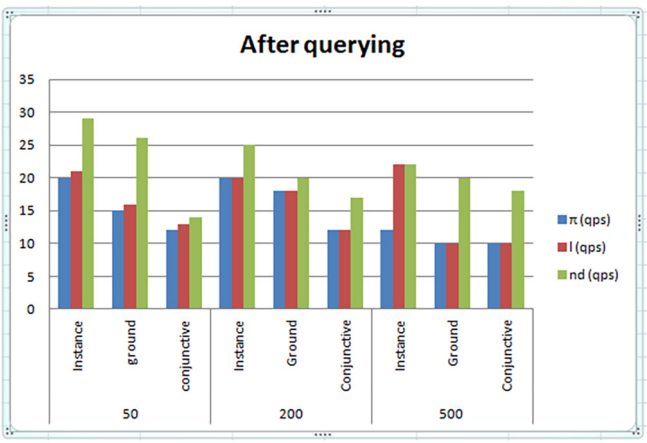


Fig. 4. Productivity of applying our repair algorithms after querying.

Definition 1. A subset $\mathcal{C} \subseteq q_{P_s}$ is said to be a conflicts sets answers if $\langle \mathcal{T}, \mathcal{C} \rangle$ is inconsistent and $\forall f \in \mathcal{C}, \langle \mathcal{T}, \mathcal{C} \setminus \{f\} \rangle$ is consistent.

4.2. Free set answers

We denote by $free(q_{P_s})$ the set of assertions belong to q_{P_s} that are not responsible for conflicts in $\langle \mathcal{T}, q_{P_s} \rangle$.

Definition 2. A *free* assertion $f \in q_{P_s}$ is said to be *free* if and only if $\forall c \in \mathcal{C}(q_{P_s}) : f \notin c$. This notion of *free* elements is formerly proposed by Benferhat et al. (1992) in a propositional logic setting.

4.3. Repairs answers

A repair answers is defined as follows:

Definition 3. A subset $\mathcal{R}_A \subseteq (q_{L_1} \cup \ldots \cup q_{L_m})$ is said to be a repair answers if $\langle \mathcal{T}, \mathcal{R}_A \rangle$ is consistent and $\mathcal{R}_A$ is said to be a maximally inclusion-based repair answers of q_{P_s}, denoted by MAR_A, if $\langle \mathcal{T}, \mathcal{R}_A \rangle$ is consistent and $\forall \mathcal{R}'_A \subseteq (q_{L_1} \cup \ldots \cup q_{L_m}) : \mathcal{R}_A \subsetneq \mathcal{R}'_A, \mathcal{R}'_A$ is inconsistent.

According to this definition of MAR_A, adding any assertion f from $(q_{L_1} \cup \ldots \cup q_{L_m}) \setminus \mathcal{R}_A$ to $\mathcal{R}_A$ implies the inconsistency of $\langle \mathcal{T}, \mathcal{R}_A \cup \{f\} \rangle$. Furthermore, the maximality in MAR_A is used in the sense of set inclusion. We denote by $MAR_A(q_{P_s})$ the set of MAR_A of q_{P_s} with respect to $\mathcal{T}$. The definition of MAR_A is similar to that defined in Lembo et al. (2010). Using the concept of repair answers, the treatment of inconsistency in flat *DL-Lite* KB s can be done by applying standard query answering either using the whole set of repairs answers (universal entailment or AR-entailment Lembo et al., 2010) or only using one repair answers (see Fig. 4).

4.4. Consistency rank

Generally, the checking of consistency degree and several inference services can be done with standard DLs reasoning services through consistent subsets of DL knowledge base has been explained in Qi et al. (2007a,b). Clearly, the computing of inconsistency degree comes down to perform a dichotomie search in standard DL, and it is closely related to the method proposed in Dubois et al. (1994) for computing inconsistency degrees of a possibilistic propositional knowledge base.

This notion of consistency rank defined is inspired by the degree of inconsistency used in the possibilistic logic where the degrees are encoded using values in the unit interval $[0, 1]$. It is counterpart of the approach proposed in Dubois et al. (1994) (*resp.* Qi et al., 2007a) in the propositional logic (*resp.* description logic) setting.

Definition 4. The consistency rank of q_{P_s}, denoted by: $CnsRank(q_{P_s})$ is defined as follows:

$$CnsRank(q_{P_s}) = max\{i \text{ where } \langle \mathcal{T}, (q_{L_1}, \ldots, q_{L_i}) \rangle \text{ is consistent}\}$$

We propose the recursive Function 1 which calculates the consistency rank of q_{P_s}. Then, we will use it in all our proposed strategies of computing a consistent assertions associated to the answers. Formally, this recursive function is faster than the sequential one (Dubois et al., 1994). The consistency rank given in Algorithm 1 is related to the notion of inconsistency degree used in possibilistic logic. The dichotomic research is made to find a Constant level of assertions where K is consistent. If n is the number of different strata in qp_s, our proposed function needs $log_2(n)$ tests. Hence, Algorithm 1 returns the rank of consistency in polynomial time.

5. Assertions associated with repairing the response profile

This section proposes three repairs to cope with inconsistent answers that seen as a set of facts. The input of these approaches is a prioritized *DL-Lite* KB with the prioritized profile, a query and the profile of sets of answers. The output of our approach is a consistent assertions associated to the sets of these answers (repair answers).

5.1. Possibilistic repair of inconsistent responses

Possibility theory (Dubois and Prade, 2007) and possibilistic logic (Dubois et al., 1994) are natural frameworks to deal with uncertain, incomplete, qualitative and prioritized information. One of the interesting aspects of possibilistic KBs is the ability of reasoning with partially inconsistent knowledge (Dubois and Prade, 1991). As shown in Benferhat et al. (2013), the entailment in possibilistic *DL-Lite*, an adaptation of *DL-Lite* entailment within a possibility theory setting, is based on the selection of one consistent, (not necessarily maximal) subset of $\mathcal{K}$. The subset $\pi(q_{P_s})$ is formed by assertions with priority levels that are less or equal to $CnsRank(q_{P_s})$.

Algorithm 1. Calculation of the Consistency Rank

Data: Inconsistent assertions associated to $\langle \mathcal{T}, (q_{L_i}, \ldots, q_{L_m}) \rangle$

Result: Consistency rank

if $\langle \mathcal{T}, (q_{L_i}, \ldots, q_{L_m}) \rangle$ *is consistent* **then**
  return m

else
  $\alpha \leftarrow i$
  $\beta \leftarrow m$
  $\gamma \leftarrow \lfloor \frac{\alpha+\beta}{2} \rfloor$
  if $\langle \mathcal{T}, (q_{L_i} \cup \ldots \cup q_{L_\gamma}) \rangle$ *is consistent* **then**
    $CnsRank(\mathcal{T}, (q_{L_\gamma}, \ldots, q_{L_m}))$
  else
    $CnsRank(\mathcal{T}, (q_{L_i}, \ldots, q_{L_\gamma}))$
  end

end

More formally, $\pi(q_{P_s})$ is the repair answers of q_{P_s} defined by $\pi(q_{P_s}) = q_{L_1} \cup \ldots \cup q_{(CnsRank(q_{P_s}))}$. If q_{P_s} is consistent with the TBox then we simply let $\pi(q_{P_s}) = q_{L_1} \cup \ldots \cup q_{L_m}$. The Algorithm 2 returns the possibilistic-based repair answers.

Algorithm 2. Possibilistic Repair of Inconsistent Responses

Data: Inconsistent assertions associated to $\langle \mathcal{T}, (q_{L_i}, \ldots, q_{L_m}) \rangle$
Result: Consistent assertions $\pi(q_{P_s})$
$\alpha \leftarrow$ *CnsRank*$(\mathcal{T}, (q_{L_1}, \ldots, q_{L_m}))$/*`Calculate recursively the consistency rank`
Return $\pi(q_{P_s}) = q_{L_1} \cup \ldots \cup q_\alpha$

This algorithm requires $\log_2(m)$ inconsistency tests on a set of m assertions associated to m answers to a query $q(x)$ with respect to P_s. It returns the possibilistic based repair answers in polynomial time.

The possibilistic conclusions are considered intact since our algorithm stops in the first assertions associated to an answer S_i where inconsistency is introduced. Hence, only the assertions having a degree strictly less or equal than the one the consistency degree are taken into account. However, assertions with priority levels strictly greater than the consistency degree are simply inhibited despite being unaffected by any conflict. To overcome this limitation, we improve possibilistic repair of inconsistent responses to linear repair of inconsistent responses.

5.2. Linear repair of inconsistent responses

In order to recover the assertions inhibited by possibilistic repair of inconsistent responses, we propose a new way corresponds to the use of linear repair of inconsistent responses from q_{P_s}.

Definition 5. Let $\mathcal{K}_P = \mathcal{T}, P_s$ be a prioritized *DL-Lite* KB with:

$P_s = \{L_1, \ldots, L_m\}, q(x)$ be a query, $S_{P_s} = \{S_1, \ldots, S_m\}$ a set of answers to a query $q(x)$ with respect to $P_s, q_{P_s} = (q_{L_1}, \ldots, q_{L_m})$ is the set of assertions associated to S_{P_s} such that $q_{L_i} = \{q(s) : s \in S_i\}$. The Linear Repair of Inconsistent Responses of q_{P_s}, denoted by: $\ell(q_{P_s}) = S'_1 \cup \ldots \cup S'_m$ is defined as follows:

$$S'_i = \begin{cases} q_{L_i} & \text{if} \langle \mathcal{T}, S'_1 \cup \ldots \cup S'_{i-1} \cup q_{L_i} \rangle \text{isconsistent} \\ \varnothing & \textit{Otherwise} \end{cases} \quad (1)$$

$\ell(q_{P_s})$ is obtained by discarding the set of assertions q_{L_i} when it conflicts with the previous set $q_{L_{i-1}}$. The following Algorithm 3 implements the $\ell(q_{P_s})$.

Algorithm 3. Linear Repair of Inconsistent Responses

Data: Inconsistent assertions associated to $\langle \mathcal{T}, (q_{L_i}, \ldots, q_{L_m}) \rangle$

Result: Consistent assertions $\pi(q_{P_s})$

$\alpha \leftarrow CnsRank(\mathcal{T}, (q_{L_1}, \ldots, q_{L_m}))$; `/* Calculate recursively the consistency rank */`

return $\pi(q_{P_s}) = q_{L_1} \cup \ldots \cup q_\alpha$

The time complexity of computing $\ell(q_{P_s})$ is in P. In fact, according to Algorithm 3, the computational complexity of computing $\ell(q_{P_s})$ needs m executions to verify the consistency of the set of assertions q_{P_s}.

5.3. Non-defeated repair of inconsistent responses

This new inference makes also to get a preferred repair answers. It consists in determining among the union of the set of assertions associated to the answers to a given query with respect to the ABox Profile P_s, the set of *free* elements.

Definition 6. Let $\mathcal{K}_P = \mathcal{T}, P_s$ be a prioritized *DL-Lite* KB with: $P_s = \{L_1, \ldots, L_m\}, q(x)$ be a query, $S_{P_s} = \{S_1, \ldots, S_m\}$ a set of answers to a query $q(x)$ with respect to $P_s, q_{P_s} = (q_{L_1}, \ldots, q_{L_m})$ is the set of assertions associated to S_{P_s} such that $q_{L_i} = \{q(s) : s \in S_i\}$. We define The *non-defeated* repair of inconsistent responses, denoted by: $nd(q_{P_s}) = S'_1 \cup \ldots \cup S'_m$ as follows:

$$\forall i = 1 \ldots m, S'_i = free(q_{L_1} \cup \ldots \cup q_{L_i})$$

Namely,
$nd(q_{P_s}) = free(q_{L_1}) \cup free(q_{L_1} \cup q_{L_1}) \cup \ldots \cup free(q_{L_1} \cup \ldots \cup q_{L_m})$.

The *non-defeated* repair is computed in polynomial time in *DL-Lite* but its computation is hard in propositional logic setting. Algorithm 4 computes the *non-defeated* repair of inconsistent responses. The complexity of this algorithm is $\mathcal{O}(m)$ where m is the number of answers to a query $q(x)$ w.r.t. the profile P_s.

Algorithm 4. *Non-defeated* Repair of Inconsistent Responses

Data: Inconsistent assertions associated to $\langle \mathcal{T}, (q_{L_1}, \ldots, q_{L_m}) \rangle$

Result: Consistent assertions $\ell(q_{P_s})$

$\alpha \leftarrow CnsRank(\mathcal{T}, (q_{L_1}, \ldots, q_{L_m}))$; `/* Calculate recursively the consistency rank */`

$\ell(q_{P_s}) \leftarrow \pi(q_{P_s})$

for $i = \alpha + 1$ *to* m **do**

  if $\langle \mathcal{T}, \ell(q_{P_s}) \cup q_{L_i} \rangle$ *is consistent* **then**

    $\ell(q_{P_s}) \leftarrow \ell(q_{P_s}) \cup q_{L_i}$

  end

end

return $\ell(q_{P_s})$

In the following example, we show that the productivity of our repairs answers strategies applied on prioritized KB after running querying is more than the repairs strategies applied on the whole KB before querying. We mean by productivity, the number of answers returned by applying each algorithm.

Example 1. Let $\mathcal{K} = \langle \mathcal{T}, P_s \rangle$ be a prioritized *DL-Lite* knowledge base. Then we have:

$\mathcal{T} = \{A \sqsubseteq \neq gB, A \sqsubseteq \neq gE, E \sqsubseteq D, R \sqsubseteq P\}$ and $P_s = \{L_1, L_2, L_3, L_4, L_5\}$ such that:

- $L_1 = \{A(a), R(a,z), A(c)\}$,
  $L_2 = \{B(a), R(b,z), A(b)\}$,
  $L_3 = \{B(a), R(a,z), B(c)\}$,
  $L_4 = \{E(e), R(e,z), A(c)\}$,
  $L_5 = \{A(e), R(e,z), A(c), R(c,z)\}$.

Let we have the following query (giving all concepts x which are on relation with z):

$q(x) = \exists x.R(x,z)$. We get the following set of assertions q_{P_s}:

- $q_{L_1} = \{A(a)\}$,
  $q_{L_2} = \{A(b)\}$,
  $q_{L_3} = \{B(a)\}$,
  $q_{L_4} = \{E(e)\}$,
  $q_{L_5} = \{A(e), A(c)\}$

One can check that the set of conflicts is answers:

- $\mathcal{C}(qp_s) = \{(A(a), B(a)); (A(e), E(e))\}$

However, the set of conflicts is of KB (according to Benferhat et al., 2015; Telli et al., 2017):

- $\mathcal{C}(P_s) = \{(A(a), B(a)); (A(e), E(e)); (A(c), B(c))\}$

In addition, the set of *free* elements for the assertion qp_s is:

- $free(qp_s) = \{A(b), A(c)\}$

While that, the set *free* elements for P_s is:

$$free(P_s) = \{R(a,z), R(b,z), A(b), R(e,z), R(c,z)\}$$

According to Algorithm 1, the consistency rank of qp_s equal 2. While, the consistency rank of P_s equal 1. Moreover, we have the following free sets of qp_s:

$$\begin{aligned} & free(q_{L_1}) = \{A(a)\}, \\ & free(q_{L_1} \cup q_{L_2}) = \{A(a), A(b)\}, \\ & free(q_{L_1} \cup \ldots \cup q_{L_3}) = \{A(a), A(b)\}, \\ & free(q_{L_1} \cup \ldots \cup q_{L_4}) = \{A(a), A(b), E(e)\}, \\ & free(q_{L_1} \cup \ldots \cup q_{L_5}) = \{A(a), A(b), E(e), A(c)\} \end{aligned}$$

Using the definitions of possibilistic repair of inconsistent responses, linear repair of inconsistent responses and *non-defeated* repair of inconsistent responses, we have:

$$\begin{aligned} & \pi(q_{P_s}) = \{A(a), A(b)\}, \\ & \ell(q_{P_s}) = \{A(a), A(b), E(e)\}, \\ & nd(q_{P_s}) = \{A(a), A(b), E(e), A(c)\} \end{aligned}$$

Now, we have the following free sets of P_s:

$$\begin{aligned} & free(L_1) = \{A(a), R(a,z), A(c)\}, \\ & free(L_1 \cup L_2) = \{A(a), R(a,z), A(c), R(b,z), A(b)\}, \\ & free(L_1 \cup \ldots \cup L_3) = \{A(a), R(a,z), A(c), R(b,z), A(b)\}, \\ & free(L_1 \cup \ldots \cup L_4) = \\ & \{A(a), R(a,z), A(c), R(b,z), A(b), E(e), R(e,z)\}, \\ & free(L_1 \cup \ldots \cup L_5) = \\ & \{A(a), R(a,z), A(c), R(b,z), A(b), E(e), R(e,z), R(c,z)\} \end{aligned}$$

Although, if we use the definitions of possibilistic repair, linear repair and *non-defeated* repair proposed in Benferhat et al. (2015) and Telli et al. (2017) directly on P_s (before querying), we have:

$$\begin{aligned} & \pi(P_s) = \{A(a), R(a,z), A(c)\}, \\ & \ell(P_s) = \{A(a), R(a,z), A(c), R(e,z), E(e)\}, \\ & nd(P_s) = \{A(a), R(a,z), A(c), R(e,z), E(e), R(c,z)\} \end{aligned}$$

By applying the same query $q(x)$ on these repairs, we have:

$$\begin{aligned} & \pi(P_{s)} \models q = \{A(a)\}, \\ & \ell(P_s) \models q = \{A(a), E(e)\}, \\ & nd(P_s) \models q = \{A(a), A(c), E(e)\}\ . \end{aligned}$$

Clearly,
$\pi(P_s) \models q \subseteq \pi(q_{P_s}), \ell(P_s) \models q \subseteq \ell(q_{P_s}), nd(P_s) \models q \subseteq nd(q_{P_s})$.

Next section is an experimental study which based on recursive programming (Consistency Rank function) about the running time of our approach.

Table 1
Statistic of the test corpus.

Used TBox	LUBM$^{\exists}$20 ontology
Total number of assertions in the ABox	1000
Number of conflicts in the ABox respectively	50, 200 and 500

6. Experimental study and results analysis

6.1. Description of test corpora

We implemented our proposed algorithms in Java for computing a repair answering in prioritized assertional bases under inconsistent KB s. The parses *DL-Lite* KB s expressed in *OWL2-QL* function syntax and a SQLite database engine. We used a part of benchmark[1] we considered the LUBM$^{\exists}$20 ontology (*i.e.*, TBox) (Lutz et al., 2013), and we generated by the Extended University Data Generator (EUDG) an ABox contains 1000 assertions and we split them into 5 strata. These ABoxes with respectively *50*, *200* and *500* conflict sets. We have executed algorithms in Telli et al. (2017) and our developed algorithms for computing repairs before and after launching an instance, ground and conjunctive query (see Table 1).

We are interested in the basic performance metrics used for evaluate our algorithms. In our case, our system classifies the assertions of the ABox into two classes: consistent and inconsistent. Consistent assertions are placed by the system in the positive class, and inconsistent assertions are placed by the system in the negative class.

We compute the four following performance indices:

- CR is the number of consistent retrieved assertions after applying the repairs.
- CNR is the number of consistent not retrieved assertions after applying the repairs.
- IR is the number of inconsistent retrieved assertions before applying the repairs.
- INR is the number of inconsistent not retrieved assertions before applying the repairs.

6.2. Evaluation measures

In the following, we present the precision, recall, F-measure, Accuracy and G-mean measures (Manning and Schiitzer, 2001; Raghavan et al., 1989), which we will use to evaluate the performance of our algorithms:

- Precision (P) is the ratio of the number of consistent assertions retrieved by the total number of retrieved assertions.
  The principle of precision measure: when we ask a query on our ABox, we wish that the assertions proposed as answer correspond to our expectations. All retrieved irrelevant assertions constitute what is called "the noise". Precision, is opposed by assertional noise. If it is high, this indicates that few unnecessary assertions are offered by the system and that the system can be considered "precise".

$$P = \frac{CR}{CR + IR} \tag{2}$$

- Recall (R) is the ratio of the number of consistent assertions retrieved by the total number of consistent assertions.
  The principle of recall measure: when we ask a query on our ABox, we wish to see appearing all assertions that could

[1] Available at:https://code.google.com/p/combo-obda/

answer our need of information. If this correspondence between the query and the number of assertions presented is important then the recall rate is high. Conversely, if the system have many interesting assertions but they do not appear in the list of answers, we speak of "silence". Silence opposes recall.

$$R = \frac{CR}{CR + CNR} \tag{3}$$

- F-measure (F) is the harmonic average of the precision P and the recall R:

$$F = \frac{2 * (P * R)}{(P + R)} \tag{4}$$

- Accuracy (Acc) is the proportion of retrieved assertions among the total number of assertions:
  Accuracy is the most intuitive performance measure, one may think that, if we have high accuracy then our model is best. Yes, accuracy is a great measure but only when you have symmetric datasets where values of consistent not retrieved and inconsistent not retrieved are almost same.

$$Accuracy = \frac{IR + CR}{INR + CR + IR + CNR} \tag{5}$$

Table 2
Diagram of confusion matrix.

	Predicted consistent assertions	Predicted inconsistent assertions
Consistent retrieved assertions	CR	INR
Inconsistent retrieved assertions	CNR	IR

- G-mean indicates the geometric mean of the recall:
  The Geometric Mean (G-Mean) is a metric that measures the balance between classification performances on both the majority and minority classes. A low G-Mean is an indication of a poor performance in the classification of the consistent assertions even if the inconsistent assertions are correctly classified (see Table 2).

$$G - mean = \sqrt{\frac{IR}{IR + CNR} \cdot \frac{CR}{CR + INR}} \tag{6}$$

6.3. Evaluations and results

The Tables 3–5 show the precision, recall, F-measure, Accuracy and G-mean measures for our algorithms, after launching an instance, ground and conjunctive query on our ABox.

All retrieved inconsistent assertions constitute what is called "the noise". Precision, is opposed by assertional noise. If it is high, this indicates that few unnecessary assertions are offered by the system and that the system can be considered "precise".

Now, we are interested in the time taken to compute our proposed algorithms of repairs. For this aim, we generated and splitted the ABoxes respectively into 3 strata, 5 strata and then 7 strata. These ABoxes contain respectively 50, 200 and 500 conflict sets. The results of this experimentation are shown in Table 6. The computational complexity analysis shows that these strategies are achieved in polynomial time.

Now, we are interested to evaluate the productivity of our repair algorithms before and after querying. We mean by productivity, the assertions that are preserved from the ABox (*resp.* answers) in order to restore the consistency of the KB (*resp.* answers).

The analysis of these results and the comparison between approaches are presented in the following section.

Table 3
Experimental evaluation of possibilistic repair of inconsistent responses expressed in %.

Conflict size	Query type	$\pi(P_s)$				
		P	R	F	Acc	G-mean
50	Instance	84.03	94.33	88.88	48.55	37.41
	Ground	79.78	92.59	85.70	52.30	48.65
	Conjunctive	75.94	90.90	82.74	54.33	51.80
200	Instance	55.55	83.33	66.66	46.55	52.82
	Ground	52.94	81.81	64.28	46.09	49.69
	Conjunctive	42.85	75	54.53	48.36	50.83
500	Instance	24.24	48.97	32.42	42.53	42.60
	Ground	21.05	44.44	28.56	44.61	42.13
	Conjunctive	21.05	44.44	28.56	47.08	42.89

Table 4
Experimental evaluation of linear repair of inconsistent responses expressed in %.

Conflict size	Query type	$\ell(P_s)$				
		P	R	F	Acc	G-mean
50	Instance	84.67	94.59	89.35	53.62	43.84
	Ground	80.80	93.02	86.48	57.69	53.62
	Conjunctive	77.38	91.54	83.86	59.84	56.93
200	Instance	55.55	83.33	66.66	48.96	57.08
	Ground	52.94	81.81	64.28	47.87	52.52
	Conjunctive	42.85	75	54.53	50.18	53.74
500	Instance	24.24	48.97	32.42	44.77	46.19
	Ground	21.05	44.44	28.56	45.53	43.69
	Conjunctive	21.05	44.44	28.56	47.87	44.27

Table 5
Experimental evaluation of non-defeated repair of inconsistent responses expressed in %.

Conflict size	Query type	$nd(P_s)$				
		P	R	F	Acc	G-mean
50	Instance	88.41	96.02	92.05	63.04	55.67
	Ground	87.24	95.58	91.21	64.61	61.50
	Conjunctive	78.65	92.10	84.84	64.56	62.03
200	Instance	60.97	86.20	71.42	50	59.05
	Ground	55.55	83.33	66.66	49.64	55.47
	Conjunctive	51.51	80.95	62.95	52.72	58.01
500	Instance	36.97	63.76	46.80	48.50	52.32
	Ground	34.78	61.53	44.44	47.23	46.51
	Conjunctive	32.43	59.01	41.85	49.13	46.45

Table 6
Runing time of our repairs before and after querying (in seconds).

Conflict size	Strata level	Before querying (Benferhat et al., 2015; Telli et al., 2017)			After querying		
		$\pi(P_s)$	$\ell(P_s)$	$nd(P_s)$	$\pi(q_{P_s})$	$\ell(q_{P_s})$	$nd(q_{P_s})$
50	3	10.93	11.09	45.45	2.83	2.88	11.50
	5	10.96	32.65	67.43	2.85	8.22	17.50
	7	16.36	32.92	94.26	4.10	8.25	23.58
200	3	12.20	28.17	47.90	3.06	3.80	11.96
	5	12.47	33.09	78.18	3.12	8.25	19.55
	7	16.53	38.53	96.99	4.17	9.65	24.25
500	3	13.86	14.41	50.68	3.45	7.07	12.77
	5	18.12	29.98	93.36	4.55	8.27	23.50
	7	19.89	43.41	99.43	4.98	10.90	25.00

6.4. Discussion

By analyzing the results obtained in the series of experiments, we notice the change in performance between the different approaches. The precision measure obtained in Tables 3–5 show that the non-defeated algorithm is more precise than the two others. Also, when increasing the number of conflicts, the precision measure decreases in the three algorithms. Hence, this measure is influenced by the number of conflicts in the ABox.

Now, if the correspondence between the query and the number of assertions presented is important, then the recall rate is high. Conversely, if the system have many interesting assertions but they do not appear in the list of answers, we speak of "silence" (silence opposes recall).

Similarly, according to the results of the Tables 3–5, we note that the recall measure of the non-defeated algorithm is higher than the other algorithms. In addition, when increasing the number of conflicts, the recall measure decreases with the three algorithms. Thus, this measure is influenced by the number of conflicts in the ABox.

Finally, according to the measures of Accuracy and G-mean calculated in Tables 3–5, we also noticed that non-defeated algorithm is the best repair algorithm compared to the others so that the values of these measures are influenced by increasing the number of conflicts.

According The results illustrated in Table 6, the existing strategies which based on access to the whole Knowledge Base take more running time than our proposed strategies which handle only with the set of answer of a specific query. Precisely, the computing of our repairs after querying requires in most of our experiments, less time than its computing before querying. However, the time needed for computing the non-defeated repair answers increases with the size of conflicts in the ABox. Finally, according to the results obtained, we conclude that non-defeated algorithm is the most performing followed by linear and possibilistic algorithms.

From Table 7, Figs. 1 and 2, the productivity of possibilistic repair of Inconsistent Responses is very cautious comparing to the other strategies. Namely, for a given ABox and a given number of strata, possibilistic repair of Inconsistent Responses has the largest number of dropped elements. This similarly holds for the linear repair of Inconsistent Responses when there exists at least a conflict in each strata. Hence, it is obvious that the size of conflicting elements in the assertional bases is one of main parameters

Table 7
Productivity of applying our repair algorithms before and after querying (expressed in %).

Conflict size	Query type	Before querying			After querying		
		$\pi(P_s) \models q$	$\ell(P_s) \models q$	$nd(P_s) \models q$	$\pi(q_{P_s})$	$\ell(q_{P_s})$	$nd(q_{P_s})$
50	Instance	15	16	25	20	21	29
	Ground	14	15	24	15	16	26
	Conjunctive	10	12	14	12	13	14
200	Instance	18	19	22	20	20	25
	Ground	17	18	20	18	18	20
	Conjunctive	10	10	15	12	12	17
500	Instance	11	11	20	12	12	22
	Ground	11	11	18	10	10	20
	Conjunctive	10	10	17	10	10	18

that influence the productivity of the repairs answers. The *non-defeated* repair of Inconsistent Responses gives a significant number of consistent answers compared to the other strategies.

Consequently, the productivity of $\pi(q_{P_s}), \ell(q_{P_s})$ and $nd(q_{P_s})$ are more productive than $\pi(P_s) \models q, \ell(P_s) \models q$ and $nd(P_s) \models q$ respectively. We note that the computing on our proposed algorithms requires a polynomial running time.

7. Conclusion

In this work, we are interested in the problem of inconsistent responses from various priority DL-Lite knowledge bases. To this end, we started by giving some notions on this type of knowledge bases and on reasoning tolerant of inconsistencies with all the answers. Then, we developed a recursive function to calculate the rank of coherence, used in the algorithms that we proposed, which allowed us to manage the inconsistencies in the set of responses. The main contribution of this work is to repair all the responses obtained instead of repairing the entire knowledge base. This strategy has saved us a lot of time compared to the times achieved by the main approaches that we have studied in the literature. The solution we proposed allowed us to reduce the complexity of calculation while increasing productivity. We have conducted experimental studies on our approach, using standard performance measures, namely precision, recall and F-measure as well as execution time. Indeed, the analysis carried out on the results found showed significant productivity and an acceptable speed of execution of our approach while remaining effective.

Our future work includes three directions. The first is to conduct a much more in-depth comparative study between these approaches in order to give academics and practitioners more amplification on how to manage inconsistencies in priority DL-Lite knowledge bases. The second consists in developing a possible extension of our approach by proceeding with a possibilistic DL-Lite extension, denoted π_l-DL-Lite, to deal with the uncertainty associated with objects, concepts and relationships using the framework of the theory of possibilities based on the notion of interval. Indeed, possibilistic logic is a formalism that allows reasoning from uncertain information represented using weighted knowledge bases potentially inconsistent. It associates with the DL-Lite axioms possibilistics a strictly positive degree in the interval [0,1]. However, in practice, it can be difficult for an expert to provide the exact degrees to associate with the axioms of the possibilistic DL-Lite. For this purpose, we propose to provide the bases of the interval-based possibilistic DL-Lite, noted π_l-DL-Lite$_{core}$ as an extension of the possibilistic DL-Lite$_{core}$ by replacing the degree applied to each axiom by an interval instead of a single value. The third direction that we will consider exploring is to deal with the missing data, within these knowledge bases, which have not been taken into account by our approach or by the approaches studied in the literature, by integrating new techniques such as deleting observations (full case analysis) or by imputing data. Missing data is an inevitable problem in the practice of statistics and expertise in general. Indeed, in the medical field for example, and more precisely in epidemiology, the performance of a system rests mainly on the treatment of missing data which remains a subject in full expansion.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Arenas, M., Bertossi, L.E., Chomicki, J., 1999. Consistent query answers in inconsistent databases, in: Proceedings of the Eighteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, May 31 - June 2, 1999, Philadelphia, Pennsylvania, USA, pp. 68–79. doi: 10.1145/303976.303983. .

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.F., (Eds.), 2010. The Description Logic Handbook: Theory, Implementation, and Applications, Cambridge University Press, paperback edition. .

Baget, J., Benferhat, S., Bouraoui, Z., Croitoru, M., Mugnier, M., Papini, O., Rocher, S., Tabia, K., 2016. A general modifier-based framework for inconsistency-tolerant query answering. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25–29, 2016, pp. 513–516. .

Baral, C., Kraus, S., Minker, J., 1991. Combining multiple knowledge bases. IEEE Trans. Knowl. Data Eng. 3 (2), 208–220. https://doi.org/10.1109/69.88001.

Benferhat, S., DuboisS, D., Pradei, H., 1992. Representing default rules in. In: Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning, pp. 673–684.

Benferhat, S., Cayrol, C., Dubois, D., Lang, J., Prade, H., 1993. Inconsistency management and prioritized syntax-based entailment. In: Proceedings of the 13th International Joint Conference on Artificial Intelligence. Chambéry, France, August 28 - September 3, 1993, pp. 640–647. .

Benferhat, S., Dubois, D., Prade, H., 1995. How to infer from inconsistent beliefs without revising? In: IJCAI, pp. 1449–1457.

Benferhat, S., Dubois, D., Prade, H., 1997. Some syntactic approaches to the handling of inconsistent knowledge bases: A comparative study part 1: The flat case. Studia Logica 58 (1), 17–45. https://doi.org/10.1023/A:1004987830832.

Benferhat, S., Bouraoui, Z., Loukil, Z., 2013. Min-based fusion of possibilistic dl-lite knowledge bases. In: 2013 IEEE/WIC/ACM International Conferences on Web Intelligence, WI 2013, Atlanta, GA, USA, November 17–20, 2013, pp. 23–28. doi: 10.1109/WI-IAT.2013.4. .

Benferhat, S., Bouraoui, Z., Tabia, K., 2015. How to select one preferred assertional-based repair from inconsistent and prioritized dl-lite knowledge bases? In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, pp. 1450–1456.

Benferhat, S., Bouraoui, Z., Croitoru, M., Papini, O., Tabia, K., 2016. Non-objection inference for inconsistency-tolerant query answering. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9–15 July 2016, pp. 3684–3690. .

Bertossi, L.E., 2011. Database Repairing and Consistent Query Answering, Synthesis Lectures on Data Management. Morgan & Claypool Publishers. https://doi.org/10.2200/S00379ED1V01Y201108DTM020.

Bienvenu, M., Rosati, R., 2013. Tractable approximations of consistent query answering for robust ontology-based data access. In: IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3–9, 2013, pp. 775–781.

Bienvenu, M., Bourgauxi, C., Goasdoue, F., 2014. Querying inconsistent description logic knowledge bases under preferred repair semantics. In: Brodley and Stone, pp. 996–1002. .

Bienvenu, M., Bourgaux, C., Goasdoué, F., 2016. Query-driven repairing of inconsistent dl-lite knowledge bases. In: IJCAI: International Joint Conference on Artificial Intelligence.

Bienvenu, M., Bourgaux, C., Goasdoué, F., 2019. Computing and explaining query answers over inconsistent dl-lite knowledge bases. J. Artif. Intell. Res. 64, 563–644.

Boughammoura, R., Omri, M.N., 2017. Querying deep web data bases without accessing to data. In: Proceedings of The 13th International Conference on Natural Computation Fuzzy Systems and Knowledge Discovery.

Boughammoura, R., Omri, M.N., Hlaoua, L., 2012. Information retrieval from deep web based on visual query interpretation. Int. J. Inform. Res. Rev. 2 (4), 45–59. https://doi.org/10.4018/ijirr.2012100104.

Boughammoura, R., Hlaoua, L., Omri, M.N., 2015. G-form: A collaborative design approach to regard deep web form as galaxy of concepts. In: Cooperative Design, Visualization and Engineering - 12th International Conference, CDVE 2015, Mallorca September 20–23, Spain Proceedings, 2015, pp. 170–174. https://doi.org/10.1007/978-3-319-24132-6_20.

Brewka, G., 1989. Preferred subtheories: An extended logical framework for default reasoning. In: Proceedings of the 11th International Joint Conference on Artificial Intelligence, pp. 1043–1048.

Chomicki, J., 2006. Consistent query answering: Five easy pieces. In: Proceedings of the 11th International Conference on Database Theory, pp. 1–17. https://doi.org/10.1007/11965893_1.

Dixit, A.A., 2019. Cavsat A system for query answering over inconsistent databases. In: Proceedings of the 2019 International Conference on Management of Data, pp. 1823–1825.

Dubois, D., Prade, H., 1991. Epistemic entrenchment and possibilistic logic. Artif. Intell. 50, 223–239.

Dubois, D., Prade, H., 2007. Possibility theory. .

Dubois, D., LANG, J., Prade, H., 1994. possibilistic logic. .

Du, J., Qi, G., Shen, Y.-D., 2013. Weight-based consistent query answering over inconsistent shiq knowledge base. Knowl. Inform. Syst. 34, 335–371.

Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., Savo, D.F., 2010. Inconsistency-tolerant semantics for description logics. In: Web Reasoning and Rule Systems -

Fourth International Conference, RR 2010, Bressanone/Brixen September 22–24, Italy, pp. 103–117. https://doi.org/10.1007/978-3-642-15918-3_9.
Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., Savo, D.F., 2015. Inconsistency-tolerant query answering in ontology-based data access. J. Web Semant. 33, 3–29. https://doi.org/10.1016/j.websem.2015.04.002.
Lutz, C., Seylan, I., Toman, D., Wolterl, F., 2013. The combined approach to obda: taming role hierarchies using filters. In: 12th International Semantic Web Conference, pp. 314–330.
Manning, C.D., Schiitzer, H. (Eds.), 2001. Foundations of Statistical Natural Language Processing. MIT Press.
Maurizio, L., 2011. Ontology-based data management, in: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 5–6. doi: 10.1145/2063576.2063582. .
Nebel, B., 1994. Base revision operations and schemes: Semantics, representation and complexity. In: ECAI, pp. 341–345. .
Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M., Rosati, R., 2008. Linking data to ontologies. J. Data Semant. 10, 133–173. https://doi.org/10.1007/978-3-540-77688-8_5.
Qi, G., Pan, J.Z., Ji, Q., 2007a. Extending description logics with uncertainty reasoning in possibilistic logic. In: European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty.
Qi, G., Pan, J.Z., J, Q., 2007. A possibilistic extension of description logics. .
Raghavan, V., Bollmann, P., Jung, G., 1989. A critical investigation of recall and precision as measures of retrieval system performance. ACM Trans. Inform. Syst. 7, 205–229.
Rescher, N., Manor, R., 1970. On inference from inconsistent premisses. Theory Decis. 1, 179–217.
Rodriguez-Muro, M., Kontchakov, R., Zakharyaschev, M., 2013. Ontology-based data access: Ontop of databases, in: The Semantic Web - ISWC 2013–12th International Semantic Web Conference, Sydney, NSW, Australia, October 21–25, 2013, Proceedings, Part I, pp. 558–573. doi: 10.1007/978-3-642-41335-3_35. .
Staworko, S., Chomicki, J., Marcinkowski, J., 2012. Prioritized repairing and consistent query answering in relational databases. Ann. Math. Artif. Intell. 64, 209–246.
Telli, A., Benferhat, S., Bourahla, M., Bouraoui, Z., Tabia, K., 2017. Polynomial algorithms for computing a single preferred assertional-based repair. Kunstliche Intelligenz 31, 15–30.
Trivela, D., Stoilos, G., Vassalos, V., 2019. Query rewriting for dl ontologies under the icar semantics. In: International Joint Conference on Rules and Reasoning. Springer, pp. 144–158.

2. Annals of Mathematics and Artificial Intelligence

Article [J2] : Query answering DL-lite knowledge bases from hidden datasets.

Auteurs : **Hamdi Ghassen,** Omri Mohamed Nazih, Benferhat Salem, Bouraoui Zied and Papini Odile

Journal : Annals of Mathematics and Artificial Intelligence

Etat : Publié

Quartile : Q3

SJR : 0.38

ISSN : 1573-7470

Overview

Annals of Mathematics and Artificial Intelligence is a scholarly journal focusing on the application of mathematical methods to diverse areas of Artificial Intelligence.

- Appeals to readers using quantitative, combinatorial, logical, algebraic and algorithmic methods in AI.
- Aims to foster new areas of applied mathematics and strengthen the scientific underpinnings of AI.
- Features collections of papers focusing on one topic, appearing in volumes or separate issues.
- Influences the development of new areas of applied mathematics and AI.

Editor-in-Chief
Martin Charles Golumbic

Impact factor
1.2 (2022)

5 year impact factor
1.1 (2022)

Submission to first decision (median)
19 days

Downloads
27,919 (2022)

Home > Annals of Mathematics and Artificial Intelligence > Article

Query answering DL-lite knowledge bases from hidden datasets

Published: 05 November 2020

Volume 89, pages 271–299, (2021) Cite this article

Download PDF

Access provided by CNUDST Consortium

Ghassen Hamdi, Mohamed Nazih Omri, Salem Benferhat, Zied Bouraoui & Odile Papini

AMAI: Your manuscript entitled Query Answering DL-Lite Knowledge Datasets

Boîte de réception ×

Martin Charles Golumbic <em@editorialmanager.com>
À Ghassen

Traduire en français ×

Ref.: Ms. No. AMAI-D-19-00087R3
Query Answering DL-Lite Knowledge Bases from Hidden Datasets
Annals of Mathematics and Artificial Intelligence

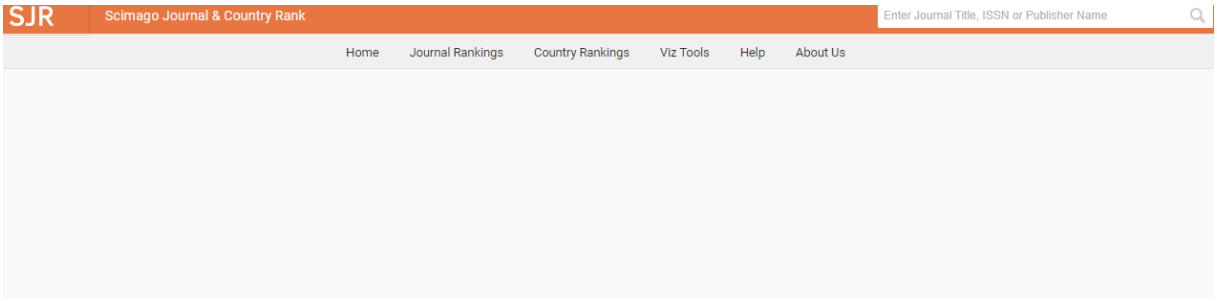
Dear Mr Hamdi,

I am pleased to tell you that your work has now been accepted for publication in Annals of Mathematics and Artificial Int

Thank you for submitting your work to this journal.

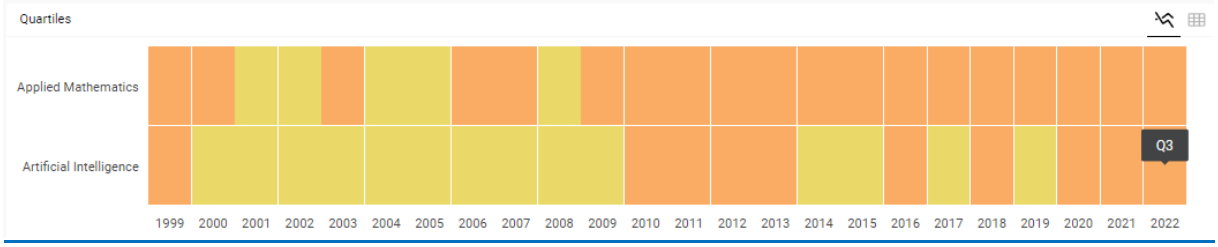
With kind regards

Martin Charles Golumbic
Editor-in-Chief
Annals of Mathematics and Artificial Intelligence



Annals of Mathematics and Artificial Intelligence

COUNTRY	SUBJECT AREA AND CATEGORY	PUBLISHER	H-INDEX
Netherlands	Computer Science └ Artificial Intelligence	Springer Netherlands	57



Annals of Mathematics and Artificial Intelligence (2021) 89:271–299
https://doi.org/10.1007/s10472-020-09714-2

Query answering DL-lite knowledge bases from hidden datasets

Ghassen Hamdi[1] · Mohamed Nazih Omri[1] · Salem Benferhat[2] · Zied Bouraoui[2] · Odile Papini[3]

Accepted: 22 September 2020 / Published online: 5 November 2020
© Springer Nature Switzerland AG 2020

Abstract
Unifying access to data using structured knowledge is the main problem studied in ontology-based data access (OBDA). Data are often provided by several information sources, and this has led to a number of methods that merge them in order to get a unified point of view. Existing merging approaches assume that the content of datasets is known and available. However, in several applications, it might be impossible to know the full content of datasets beforehand. This paper investigates several query answering strategies from multiple datasets without knowing their content in advance. We study those strategies from different points of view: productivity, logical properties and computational complexity.

Keywords Query answering · OBDA · DL-Lite · Ontology

Mathematics Subject Classification (2010) 14Q99 · 13P10 · 68T15

✉ Ghassen Hamdi
hamdighassan@gmail.com

Mohamed Nazih Omri
mohamednazih.omri@fsm.rnu.tn

Salem Benferhat
benferhat@cril.univ-artois.fr

Zied Bouraoui
bouraoui@cril.univ-artois.fr

Odile Papini
odile.papini@univ-amu.fr

1 MARS Research Laboratory, University of Sousse, Sousse, Tunisia

2 CRIL. University of Artois, Lens, France

3 Aix-Marseille Université, Université de Toulon, LIS, CNRS, Marseille, France

1 Introduction

A Knowledge Base (KB) gathers a set of knowledge of a given domain. Ontologies offer a powerful framework to represent concepts and relationships between concepts. They are typically expressed using Description Logics (DLs) [1], and stored in two components: a terminological base, called TBox, which describes the generic knowledge about a given domain, and an assertional base, called ABox, which contains the assertions (i.e., facts) that instantiate the terminological knowledge. In the recent years, Ontology-Based Data Access (OBDA) (e.g. [2–4]) have gained popularity due to their capabilites for managing information systems through semantic technologies. In this setting, structured knowledge (i.e., stored in a TBox) is exploited during querying data (i.e., stored in an ABox). The most important service provided by OBDA is query answering.

In many cases, data are provided by multiple sources, and developping meaningful query answering systems is an important research question. Several problems are related to the context of multiple sources, such as: redundancy, erroneous data, contradiction and the dissatisfaction of sources among others. To address these problems, some work has been suggested to merge data (i.e., ABoxes) in order to get a unified point of view of ABoxes (e.g. [5–7]). In order to deal with conflicts during the merging process, existing merging operators combine information derived from several sources by following several strategies that are based on the assumption that the content of sources is known in advance. However, this assumption does not always hold and data may not be accessible. For example, data might be is hidden for privacy or security reasons.

This paper studies query answering without merging data sources and without knowing their contents. We put forward several query answering strategies from hidden datasets when KBs are described in DL-Lite logics. We focus on DL-Lite logic, [8] which offers a good compromise between expressive power and computational complexity and provides an adequate framework for an efficient query answering. We first compare those strategies from a productivity point of view. We then study how the proposed strategies behave with respect to the logical properties suggested in [9] concerning the conjunction and consistency of queries. Query answering strategies may be questionable in the presence of new assertions, and the proposed strategies can be seen as non-monotonic inferences. We characterize the behavior of the suggested strategies with respect to the well-known Kraus Lehmann and Magidor rationality properties [10] rephrased within the DL-Lite framework. This paper also provides a computational complexity analysis that shows that the proposed strategies are achieved in polynomial time.

This paper is organized as follows. Section 2 recalls some basic notions of DL-Lite and query answering. Section 3 proposes several strategies to answer queries from datasets without knowing their contents. Section 4 compares the suggested strategies from a productivity point of view. Section 5 investigates the rational properties of the proposed stategies. We present, in Section 6, some related work and, in Section 7 a discussion on how to go beyond our inference strategies. Section 8 concludes this paper and highlight ongoing work.

2 Background

Description Logics (DLs) are a family of formalisms designed to represent the knowledge of a particular domain. Recently, a lot of attention has been given to DL-Lite, a family of lightweight DLs specifically designed for applications that use huge volumes of data, such as the Web applications where query answering is the most important reasoning task [11].

In particular, DL-Lite guarantees an efficient computational complexity of the reasoning process. This makes DL-Lite more suited to the context of OBDA. In the following, we recall some basic notions of a class of DL-Lite, called DL-Lite$_R$ logic [8].

2.1 DL-Lite syntax

Let N_C, N_R and N_I be three pairwise disjoint sets of atomic concepts, atomic roles and individuals respectively. Let $A \in N_C$, $P \in N_R$, and let '$\neg$', '$\exists$' and '$^-$' be three connectors of the DL-Lite language, '$\neg$' stands for set complements of (concepts or roles), '$^-$' stands for inverse (applied to roles) and '$\exists$' applied to roles, give the concept obtained from the projection of first elements of terms in roles. The language of DL-Lite is defined as follows:

$$\begin{array}{lllll} B \longrightarrow & A & | & \exists R & \quad C \longrightarrow \quad B \quad | \quad \neg B \\ R \longrightarrow & P & | & P^- & \quad E \longrightarrow \quad R \quad | \quad \neg R \end{array}$$

where A is an atomic concept, P is an atomic role and P^- is the inverse of an atomic role, B (*resp.* C) is the set of basic (*resp.* complex) concepts, and R (*resp.* E) is the set of basic (*resp.*complex) roles.

A DL-Lite KB is a pair $\mathcal{K}$=$\langle \mathcal{T}, \mathcal{A} \rangle$ where $\mathcal{T}$ is a TBox and $\mathcal{A}$ is an ABox. A DL-Lite TBox consists of a finite set of inclusion axioms of the following form:

$B \sqsubseteq C$ or $R \sqsubseteq E$.

A DL-Lite ABox consists of a finite set of membership assertions on atomic concepts and atomic roles, respectively, of the following form:

$A(a)$ and $P(a, b)$,

where $A \in N_C$, $P \in N_R$ and $a, b \in N_I$.

For the sake of simplicity, in the rest of this paper, when there is no ambiguity, we simply use DL-Lite instead of DL-Lite$_R$.

2.2 DL-Lite semantics

The semantics of DL-Lite is given in terms of interpretations. Interpretation I=$(\Delta^I, .^I)$ consists of a non-empty domain Δ^I and an interpretation function $.^I$. Function $.^I$ assigns to each individual $a \in N_I$ an element $a^I \in \Delta^I$, to each concept $C \in N_C$ a subset $C^I \subseteq \Delta^I$ and to each role $R \in N_R$ a binary relation $R^I \subseteq \Delta^I \times \Delta^I$. An interpretation function $.^I$ is extended to all constructs of DL-Lite as follows:

$$\begin{aligned} A^I &\subseteq \Delta^I, \\ (P)^I &\subseteq \Delta^I \times \Delta^I, \\ (P^-)^I &= \{(y, x) \in \Delta^I \times \Delta^I | (x, y) \in P^I\}, \\ (\exists R)^I &= \{x \in \Delta^I | \exists y \in \Delta^I \; such\, that \; (x, y) \in R^I\}, \\ (\neg B)^I &= \Delta^I \setminus B^I, \\ (\neg R)^I &= \Delta^I \times \Delta^I \setminus R^I. \end{aligned}$$

For the TBox, we say that I satisfies a concept (*resp.* role) inclusion axiom, denoted by $I \models B \sqsubseteq C$ (*resp.* $I \models R \sqsubseteq E$), if and only if $B^I \subseteq C^I$ (*resp.* $R^I \subseteq E^I$). For the ABox, we say that I satisfies a concept (*resp.* role) membership assertion, denoted by $I \models A(a_i)$ (*resp.* $I \models P(a_i, a_j)$), if and only if $a_i^I \in A^I$ (*resp.* $(a_i^I, a_j^I) \in P^I$). Note that we only consider *DL-Lite* with a unique name assumption.

An interpretation I is said to satisfy a KB $\mathcal{K}$=$\langle \mathcal{T}, \mathcal{A} \rangle$ if and only if I satisfies every axiom

in $\mathcal{T}$ and every axiom in $\mathcal{A}$. Such interpretation is said to be a model of $\mathcal{K}$ [8]. $\langle \mathcal{T}, \mathcal{A} \rangle \models \langle \mathcal{T}, \mathcal{A}' \rangle$ means that the models of $\langle \mathcal{T}, \mathcal{A} \rangle$ are models of $\langle \mathcal{T}, \mathcal{A}' \rangle$, and $\langle \mathcal{T}, \mathcal{A} \rangle \equiv \langle \mathcal{T}, \mathcal{A}' \rangle$ means that $\langle \mathcal{T}, \mathcal{A} \rangle$ and $\langle \mathcal{T}, \mathcal{A}' \rangle$ have the same models.

2.3 Incoherence and inconsistency

Two notions of conflicting knowledge can be distinguished in DL KBs: *incoherence* and *inconsistency* [1, 12]. The former is considered as a kind of inconsistency in the TBox. The latter is the classical inconsistency notion for KBs. Namely, a KB is said to be inconsistent if and only if it does not admit any model, and it is said to be incoherent if there exists at least a non-satisfiable concept (i.e. no individual can belong to the concept) or a non-satisfiable role. More formally:

Definition 1 A *DL-Lite* TBox $\mathcal{T}$ is said to be incoherent if there exists a concept (*resp.* a role) C such that for each interpretation I, which is a model of $\mathcal{T}$, we have $C^I = \emptyset$ (*resp.* $R^I = \emptyset$).

An example of an incoherent TBox is the one composed of the two inclusion axioms $\mathcal{T} = \{B_1 \sqsubseteq B_2, B_1 \sqsubseteq \neg B_2\}$. In a propositional setting, the counterpart of incoherence is called potential inconsistency, as defined for instance in [13]. KB inconsistency is defined as follows:

Definition 2 A *DL-Lite* KB $\mathcal{K}$ is said to be inconsistent if it does not admit any model.

Inconsistency in DL-Lite is always defined with respect to some ABoxes, since a TBox may be incoherent but never inconsistent, as stated by Calvanese et al. [14]. Moreover, the latter authors provided the following lemma:

Lemma 1 ([14]) *Let $\mathcal{K}$=$\langle \mathcal{T}, \mathcal{A} \rangle$ be a DL-Lite KB. If $\mathcal{A} = \emptyset$, then $\mathcal{K}$ is consistent. If $\mathcal{K}$ is inconsistent, then there exists subset $\mathcal{A}_0 \subseteq \mathcal{A}$ with two elements at most, such that $\langle \mathcal{T}, \mathcal{A}_0 \rangle$ is inconsistent.*

2.4 Query answering

A query is a first-order logic formula of the following form:

$$q(\overrightarrow{x}) = \{\overrightarrow{x} \mid \phi(\overrightarrow{x})\}$$

where $\overrightarrow{x}$=$(x_1, ..., x_n)$ are free variables, n is the arity of $q(\overrightarrow{x})$, and atoms of $\phi(\overrightarrow{x})$ are of the form $A(t_i)$ or $P(t_i, t_j)$, where $A \in N_C$, $P \in N_R$, and t_i and t_j are terms, i.e., constants or variables of N_I. In this paper, we focus on Conjunctive Queries (CQs) and Boolean Queries (BQs).

A CQ is a query of the following form:

$$q(\overrightarrow{x}) = \{\overrightarrow{x} \mid \exists \overrightarrow{y}.conj(\overrightarrow{x}, \overrightarrow{y})\}$$

where $\overrightarrow{x}$=$(x_1, ..., x_n)$ are free variables, n is the arity of $q(\overrightarrow{x})$, $\overrightarrow{y}$ are bounded variables called existentially quantified variables, and $conj(\overrightarrow{x}, \overrightarrow{y})$ is a conjunction of atoms in the form of $A(t_i)$ or $P(t_i, t_j)$, where $A \in N_C$, $P \in N_R$ and t_i and t_j are terms.

A BQ, denoted by $q()$, is a query where $n = 0$. A BQ with no bounded variables is called a grounded query. A BQ consisting of only one atom is called an instance query (also called instance checking).

Given BQ $q()$ and KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, an answer to $q()$ w.r.t. $\mathcal{K}$ is a boolean value (true or false).

The paper focuses on CQs and BQs. Let $\mathcal{A}$ be an ABox. For the sake of simplicity, we use in the following notation $\wedge f$ for the conjunction of assertions $f_1 \wedge \cdots \wedge f_n$ with $f_1, \cdots, f_n \in \mathcal{A}$, and we consider BQs of the $q() = \wedge f$ form.

3 Multiple-source query answering strategies

3.1 Context of this paper

As mentioned in the introduction section, the context of our work is the one of querying multiple-source KBs. We use lightweight ontologies, encoded by DL-Lite language, to represent and query knowledge bases.

The input of our problem is a terminological KB, denoted by $\mathcal{T}$, and a set of assertional bases, denoted by $\{\mathcal{A}_1, ..., \mathcal{A}_n\}$. The terminological base $\mathcal{T}$ is assumed to be coherent. We also assume that axioms in $\mathcal{T}$ are reliable, so they cannot be questionable in the presence of conflicts. Lastly, we assume that the terminological base is available.
As regards to the assertional base, we assume that each ABox $\mathcal{A}_i$ is individually consistent with $\mathcal{T}$. However, it is unlikely that the union of assertional bases $\mathcal{A}_1 \cup \cdots \cup \mathcal{A}_n$ is consistent with $\mathcal{T}$.

In the following, the input knowledge base will be simply denoted by:

$$\mathcal{K} = \langle \mathcal{T}, \{\mathcal{A}_1, \cdots, \mathcal{A}_n\} \rangle$$

Query answering the knowledge $\mathcal{K}$ depends on different underlying assumptions on $\mathcal{K}$, whether $\mathcal{K}$ is consistent or not and whether the ABoxes $\mathcal{A}_i{}'$s are available or not.

Figure 1 summarizes the different attitudes to query $\mathcal{K}$. These attitudes depend on the following cases :

Case 1 : The set of ABoxes $\mathcal{A}_i$ is available. This case is widely handled in the literature and is not discussed in this paper. Availability here means that the full content of each ABoxes is accessible and if needed it can be changed or even replaced by new elements. In this case, one may consider three situations.

- Case 1.1 : The knowledge base $< \mathcal{T}, \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_n >$ composed of the TBox and the set of all ABoxes, is consistent. In this case, querying $\mathcal{K} = \langle \mathcal{T}, \{\mathcal{A}_1, \cdots, \mathcal{A}_n\} \rangle$ simply comes down to use a standard querying of DL-Lite over $< \mathcal{T}, \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_n >$. This is depicted by number (1) in Fig. 1.
- Case 1.2 : The knowledge base $< \mathcal{T}, \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_n >$ is inconsistent. In this case, there are also two strategies to consider. Either one may use inconsistency-tolerant query answering over $< \mathcal{T}, \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_n >$ (e.g. AR, IAR, CAR semantics [38, 39]) or one may merge the set of ABoxes $\{\mathcal{A}_1, \cdots, \mathcal{A}_n\}$ into a consistent assertional base $\Delta(\{\mathcal{A}_1, \cdots, \mathcal{A}_n\})$ with $\mathcal{T}$ (for instance using Hamming or Dalal's distance). These two situations, depicted by numbers (2) and (3) in Fig. 1, have been largely addressed in the literature (see for instance [9, 42] for an overview).

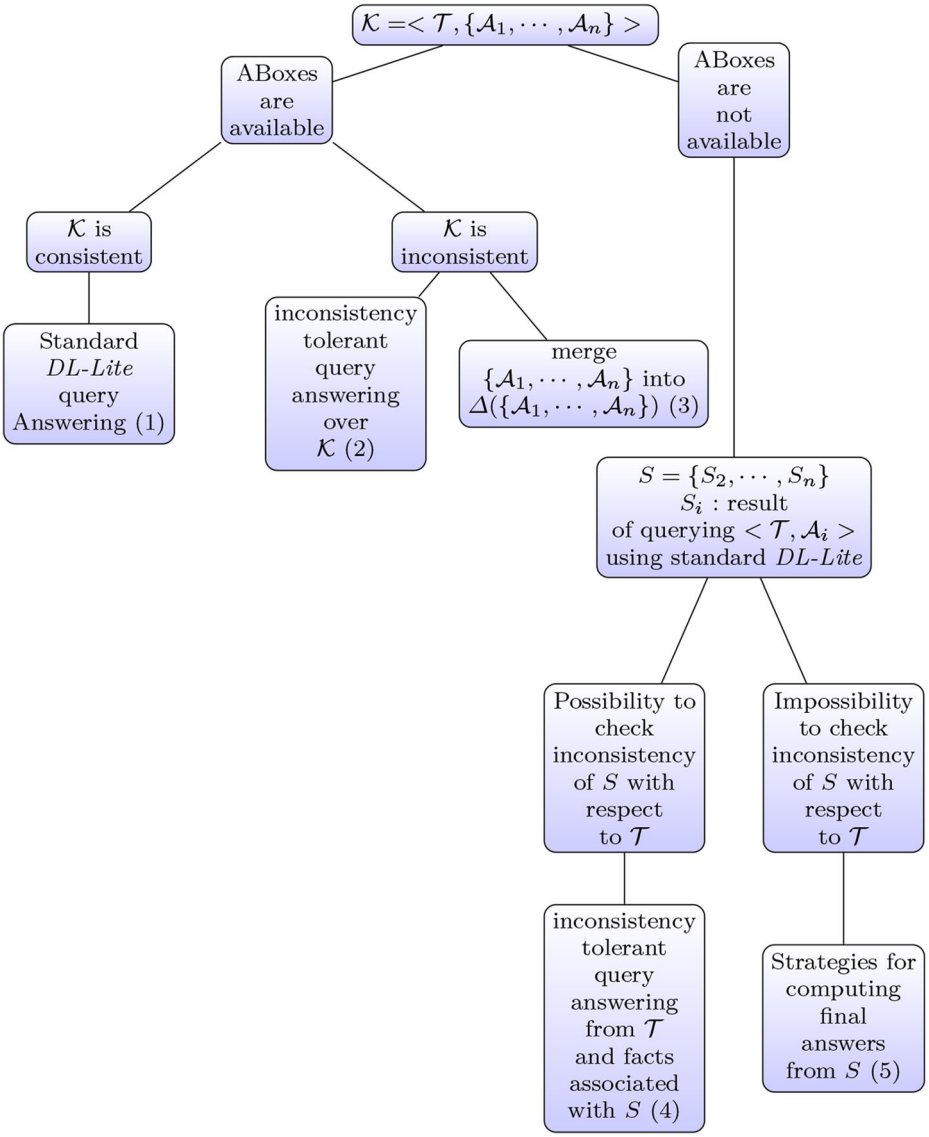


Fig. 1 Strategies for querying a set of ABoxes

Case 2 : The ABoxes $\mathcal{A}_i$ are not available. We also use the term "hidden" to describe this situation. This is a difficult situation that is addressed in the paper. In this case, ABoxes are not fully available. However, they can be individually queried. Of course, querying ABoxes should be limited and should not allow recovering full ABoxes (otherwise, they will come down to Case 1).

The first step in this case is to collect answers to a given query q from each ABox $\mathcal{A}_i$. We will denote by S_i such a set of answers.

Again, there are two strategies to consider, respectively represented by number (4) and number (5) (see Fig. 1). The first strategy (4) supposes that there is a way to check whether assertions present in each $S = \{S_1, \cdots, S_n\}$ are conflictual or not. If it is the case, then one needs to apply again inconsistency-tolerant query answering in the spirit of Case 1. This consists in analyzing different strategies to come up with a final answer to a query q from $S = \{S_1, \cdots, S_n\}$.

This paper focuses on case 2 of Fig. 1. In the next section, we define several query answering strategies in the presence of several "hidden" ABoxes.

3.2 Query answering strategies

In the following, we will use the term "profile", to denote a multiset of ABoxes. Namely, $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ is a KB where $\mathcal{T}$ is a DL-Lite TBox and $P = \{\mathcal{A}_1, ..., \mathcal{A}_m\}$ is an ABox profile. We assume that each ABox $A_i \in P$ is consistent w.r.t. $\mathcal{T}$, namely $\langle \mathcal{T}, \mathcal{A}_i \rangle$ is consistent.

Example 1 Let us consider a KB $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$. Let PhdStudents, Professors and Students be three concepts, and TutorOf and TutorialTo be two roles. Intuitively, the relation TutorOf (x,y) simply means that y has x as a tutor. Similarly, TutorialTo (x,y) means that y provides tutorials to x.

Assume that the TBox T is encoded as follow:

$$\begin{array}{l} PhdStudents \sqsubseteq \neg\, Professors \\ PhdStudents \sqsubseteq Students \\ Students \sqsubseteq \exists\, TutorOf \\ TutorOf \sqsubseteq TutorialTo \end{array}$$

The first axiom means that PhdStudents are not professors. The second axiom means that PhdStudents are students. The third axiom means that if x is student, then x has a tutor. The last axiom means that if x is a tutor of y, then x provides tutorials to y.

Let $D = \{a, b, c, d, e, f, p, q\}$ a set of individuals. Assume that we have four universities. The following four ABoxes encode the factual role played by each individual in the four universities.

The ABoxes are described as follows:

$\mathcal{A}_1$ =	{PhdStudents(a), PhdStudents(b), PhdStudents(d), Students(e), TutorOf(p, a), TutorOf(p, b), TutorOf(p, d), TutorOf(q, e), TutorialTo(p, b)},
$\mathcal{A}_2$ =	{PhdStudents(c), Professors(a), Students(b), Students(d), Students(f), TutorOf(p, a), TutorOf(p, c), TutorOf(p, d), TutorOf(q, f), TutorialTo(p, b)},
$\mathcal{A}_3$ =	{Professors(a), Professors(c), Students(b), Students(d), Students(e), Students(f), TutorOf(p, a), TutorOf(p, b), TutorOf(p, d), TutorOf(q, e), TutorOf(q, f), TutorialTo(p, b)},
$\mathcal{A}_4$ =	{PhdStudents(a), PhdStudents(b), PhdStudents(c), TutorOf(p, a), TutorOf(p, b), TutorOf(p, c), TutorialTo(q, b)}.

One can easily check that $\forall i, 1 \leq i \leq 4$, $\langle \mathcal{T}, \mathcal{A}_i \rangle$ is consistent but $< \mathcal{T}, \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4 >$ is inconsistent.

This example will be used in the rest of this paper in order to illustrate the main introduced concepts.

Let $q(\overrightarrow{x})$ be a CQ. We denote by S_i the set of answers to $q(\overrightarrow{x})$ obtained from $\mathcal{A}_i$. Using standard *DL-Lite*, namely $S_i = \{\overrightarrow{s} \in N^I \times \cdots \times N^I : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q(\overrightarrow{s})\}$. We denote by $S_P = \{S_1, ..., S_m\}$ the answers profile, which is a multiple sets of answers to query $q(\overrightarrow{x})$ with respect to P.

Intuitively, S_P contains all the answers that can be obtained from each ABox $\mathcal{A}_i \in P$.

When there is no answer to query $q(\overrightarrow{x})$, w.r.t. $\mathcal{A}_i$, we simply write $S_i = \emptyset$.

We now analyze different strategies for querying $\mathcal{K}_P$ on the basis of S_P.

One way to perform query answering is to only consider answers to the queries that holds in all the answer sets S_i. More formally :

Definition 3 Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a KB with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$. Let $q(\overrightarrow{x})$ be a conjunctive query and S_P be the profile of sets of answers to $q(\overrightarrow{x})$ w.r.t. P. An answer to query $q(\overrightarrow{x})$, w.r.t. $\mathcal{K}_P$, is a tuple $\overrightarrow{s}$ such that $\forall S_i \in S_P$, $\overrightarrow{s} \in S_i$. In this case, $\overrightarrow{s}$ is called a universal consequence of $\mathcal{K}_P$, and it is denoted by $\mathcal{K}_P \models_\forall q(\overrightarrow{s})$.

Note that the notion of a universal inference offers a natural way to deal with conflicting sources. It has been used, for instance, in default reasoning [16], where one only accepts conclusions derived from each extension of a default theory. We denote by $S_\forall$ the set of answers obtained by a universal inference, with $S_\forall = \{\overrightarrow{s} : \mathcal{K}_P \models_\forall q(\overrightarrow{s})\}$.

Example 2 Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be the DL-Lite KB introduced in Example 1. Let us analyse two different queries.

- Assume that we are interested in computing the list of students that are tutors. Namely, let $q_1(x) = \exists y\, TutorOf(y, x) \wedge Students(x)$. Let $S_P = \{S_1, S_2, S_3, S_4\}$ be the answers profile to the query $q_1(x)$ with:

$S_1 = \{a, b, d, e\}$ set of answers from ABox $\mathcal{A}_1$
$S_2 = \{b, c, d, f\}$ set of answers from ABox $\mathcal{A}_2$
$S_3 = \{b, d, e, f\}$ set of answers from ABox $\mathcal{A}_3$
$S_4 = \{a, b, c\}$ set of answers from ABox $\mathcal{A}_4$

  Clearly, $\mathcal{K}_P \models_\forall q_1(b)$ since $\forall S_i \in S_P$ $b \in S_i$ and $S_\forall = \{b\}$.
- Consider now $q_2(x) = \exists y\, TutorOf(y, x) \wedge PhdStudents(x)$ (namely, we are interested in computing the list of PhD students that are tutors). We have $S_P = \{S_1, S_2, S_3, S_4\}$ which are the answers profile to the query q_2.

$S_1 = \{a, b, d\}$ set of answers from ABox $\mathcal{A}_1$
$S_2 = \{c\}$ set of answers from ABox $\mathcal{A}_2$
$S_3 = \emptyset$ set of answers from ABox $\mathcal{A}_3$
$S_4 = \{a, b, c\}$ set of answers from ABox $\mathcal{A}_4$

  Clearly $S_\forall = \emptyset$ since $S_3 = \emptyset$.

As a second strategy, we propose the existential inference that states that an answer is deduced from at least one ABox. More formally:

Definition 4 Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a knowledge base with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$. Let $q(\overrightarrow{x})$ be a conjunctive query and S_P be the answers profile to $q(\overrightarrow{x})$ w.r.t. P. An answer $\overrightarrow{s}$ is said to be an *existential* answer to query $q(\overrightarrow{x})$ w.r.t. $\mathcal{K}_P$, denoted by $\mathcal{K}_P \models_\exists q(\overrightarrow{s})$, if there exists $S_i \in S_P$ such that $\overrightarrow{s} \in S_i$. Namely:

$$\mathcal{K}_P \models_\exists q(\overrightarrow{s}) \quad if \quad \exists S_i \in S_P, \ \overrightarrow{s} \in S_i$$

We denote by $S_\exists$ the set of answers obtained by an existential inference. Namely $S_\exists = \{\overrightarrow{s} \ : \ \mathcal{K}_P \models_\exists q(\overrightarrow{s})\}$.

Example 3 Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be the KB of Example 1. Consider again the same two queries of example 2:

- With $q_1(x) = \exists y \, TutorOf(y, x) \wedge Students(x)$. We have $\mathcal{K}_P \models_\exists q_1(a)$ since $a \in S_1$. Moreover, $\mathcal{K}_P \models_\exists q_1(b)$, $\mathcal{K}_P \models_\exists q_1(c)$, $\mathcal{K}_P \models_\exists q_1(d)$, $\mathcal{K}_P \models_\exists q_1(e)$ and $\mathcal{K}_P \models_\exists q_1(f)$. So, $S_\exists = \{a, b, c, d, e, f\}$.
- With $q_2(x) = \exists y \, TutorOf(y, x) \wedge PhdStudents(x)$. We have, $\mathcal{K}_P \models_\exists q_2(a)$ since $a \in S_1$. In addition, $\mathcal{K}_P \models_\exists q_2(b)$, $\mathcal{K}_P \models_\exists q_2(c)$, $\mathcal{K}_P \models_\exists q_2(d)$. Therefore, $S_\exists = \{a, b, c, d\}$.

The existential query-answering strategy is a very adventurous strategy. To this end, we propose to consider a majority-based strategy as an alternative strategy. It states that an answer to a query is considered as valid if it is deduced from the majority of datasets. More formally:

Definition 5 Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a KB with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$. Let $q(\overrightarrow{x})$ be a CQ and S_P be the answers profile to $q(\overrightarrow{x})$ w.r.t. P. An answer $\overrightarrow{x}$ is said to be a *majority* answer to query $q(\overrightarrow{x})$ w.r.t. $\mathcal{K}_P$, denoted by $\mathcal{K}_P \models_{maj} q(\overrightarrow{s})$, if $\frac{|\{i \ : \ \overrightarrow{s} \in S_i\}|}{m} > \frac{1}{2}$. Namely:

$$\mathcal{K}_P \models_{maj} q(\overrightarrow{s}) \quad if \quad \frac{|\ \{i \ : \ \overrightarrow{s} \in S_i\}\ |}{m} > \frac{1}{2}.$$

We denote by S_{maj} the set of answers obtained by the majority-based strategy, namely $S_{maj} = \{\overrightarrow{s} \ : \ \mathcal{K}_P \models_{maj} q(\overrightarrow{s})\}$. Note that this strategy offers a good compromise between the universal and existential strategies.

Example 4 Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be the KB of Example 1. Consider again the query $q_1(x)$ of example 2, namely :

$q_1(x) = \exists y \, TutorOf(y, x) \wedge Students(x)$. One can check that $\mathcal{K}_P \models_{maj} q_1(b)$, since $\forall S_i, b \in S_i$. Besides, $\mathcal{K}_P \models_{maj} q_1(d)$ since $d \in S_1$, $d \in S_2$ and $d \in S_3$. Hence, $S_{maj} = \{b, d\}$.

3.3 Introducing cardinality

We now introduce strategies obtained by adding a cardinality criterion to the aforementioned strategies. We basically restrict the application of the query answering relations $\forall$, $\exists$ and maj to some sets of answers in profile S_p. More precisely, we only select the largest sets in S_P with respect to cardinality.

We denote by Sc_P the profile of sets of answers with maximal cardinality to a query $q(\overrightarrow{x})$ w.r.t. P. Sc_P is defined as follows:

$$Sc_P = \{S_i,\ i \in \{1, \cdots, m\}\ :\ \forall j \in \{1, \cdots, m\}\ \mid S_j \mid \leq \mid S_i \mid\}$$

The strategies presented before can be extended using the cardinality criterion as follows:

Definition 6 Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a KB with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$. Let $q(\overrightarrow{x})$ be a CQ and Sc_P be the answers profile with maximal cardinality to $q(\overrightarrow{x})$ w.r.t. P. An $\overrightarrow{s}$ answer to CQ $q(\overrightarrow{x})$, w.r.t. $\mathcal{K}_P$, is said to :

- Card-universal answer: $\mathcal{K}_P \models_{\forall_c} q(\overrightarrow{s})$ if $\forall S_i \in Sc_P, \overrightarrow{s} \in S_i$.
- Card-existential answer: $\mathcal{K}_P \models_{\exists_c} q(\overrightarrow{s})$ if $\exists S_i \in Sc_P, \overrightarrow{s} \in S_i$.
- Card-majority answer: $\mathcal{K}_P \models_{maj_c} q(\overrightarrow{s})$ if $\frac{|\{S_i \in Sc_P\ :\ \overrightarrow{s} \in S_i\}|}{|Sc_P|} > \frac{1}{2}$.

We denote by $S_{\forall_c}$ (resp. $S_{\exists_c}$, S_{maj_c}) the set of answers obtained by universal (resp. existential, majority-based) strategies over Sc_P.

Example 5 Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be the KB from Example 1. Consider again the same two queries of Example 2:

- With $q_1(x) = \exists y\, TutorOf(y, x) \wedge Students(x)$. Note first that $Sc_P = \{S_1, S_2, S_3\}$. We have $S_{\forall_c} = \{b, d\}$, which expresses the fact that individuals b and d appear in all the answers profile with maximal cardinality. $S_{\exists_c} = \{a, b, c, d, e, f\}$ indicates that individuals a, b, c, d, e and f appear at least in one of the answers profile with maximal cardinality. $S_{maj_c} = \{b, d, e, f\}$ shows that individuals b, d, e and f appear in the majority of the answers profile with maximal cardinality.
- With $q_2(x) = \exists y\, TutorOf(y, x) \wedge PhdStudents(x)$. We have, $Sc_P = \{S_1, S_4\}$. Hence $S_{\forall_c} = \{a, b\}$, which expresses the fact that individuals a and b appear in all the answers profile with maximal cardinality. $S_{\exists_c} = \{a, b, c, d\}$ and $S_{maj_c} = \{a, b\}$.

3.4 Boolean queries

We now consider BQs, a special case of CQs.

Definition 7 Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a KB with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$ and $q()$ be a BQ. An answer to the query $q()$, w.r.t. $\mathcal{A}_i$, is a Boolean value b_i such that $b_i = true$ if $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q()$ and $b_i = false$ otherwise.

In order to present the inferences introduced above for BQs, we denote by $B_P = \{b_1, \cdots, b_m\}$ the set of answers profile to a BQ $q()$ w.r.t. $\mathcal{K}_P$. The following rephrases the definitions of the above strategies when applied to boolean queries:

Definition 8 Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a KB with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$. Let $q()$ be a BQ and B_P be the set of answers to $q()$ w.r.t. P.

– A universal answer to a BQ $q()$, w.r.t. $\mathcal{K}_P$, is as follows:
$\mathcal{K}_P \models_{\forall} q()$ if $\forall b_i \in B_P,\ b_i = true$, in other words $\forall \mathcal{A}_i \in P,\ \langle \mathcal{T}, \mathcal{A}_i \rangle \models q()$.

– An existential answer to a BQ $q()$, w.r.t. $\mathcal{K}_P$, is such that:
$\mathcal{K}_P \models_{\exists} q()$ if $\exists b_i \in B_P,\ b_i = true$, i.e., $\exists \mathcal{A}_i \in P,\ \langle \mathcal{T}, \mathcal{A}_i \rangle \models q()$.

– A majority answer to a BQ $q()$, w.r.t. $\mathcal{K}_P$, is:
$\mathcal{K}_P \models_{maj} q()$ if $\frac{|\{i:\ b_i = True\}|}{m} > \frac{1}{2}$, that is to say $\frac{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q()\}|}{m} > \frac{1}{2}$.

Example 6 Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$, with $P = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_3\}$, be the KB from Example 1. Consider BQ $q_1() = TutorOf(p, a)$. Clearly $\mathcal{K}_P \models_{\forall} q_1()$. Now consider BQ $q_2() = TutorOf(q, e)$. We have $\mathcal{K}_P \models_{\exists} q_2()$. Finally, consider BQ $q_3() = TutorOf(p, b)$. We have $\mathcal{K}_P \models_{maj} q_3()$.

In the special case of BQs (where there is only one answer to the query), cardinality is not very relevant.

Before studying the productivity relation and rational properties of query answering strategies, let us first briefly provides their computational complexity from a DL-Lite knowledge base without knowing the contents of the datasets. We assume that the TBox is fixed and that the size of each dataset is bounded by a constant as well as the size of the queries. We consider the studied complexity problems under data complexity as it is usually the case for OBDA [8].

We first recall the complexity result of (CQ) answering in DL-Lite.

The CQ answering problem is in AC^0 for data complexity [11] is standard *DL-Lite*. In our approach, since we assume that the datasets are not accessible, the proposed query answering strategies rely on getting all answers from each ABox. This is an enumeration problem which is out of the scope of our paper. While standard DL-Lite knowledge bases have a single ABox, the suggested strategies of query answering generalize conjunctive query answering from a knowledge base with a multiset of ABoxes, namely $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$. Once we get all answers to the query from the m datasets, namely $S_1, \ldots, S_m$ in order to evaluate the complexity of the proposed query answering strategies, we consider the following decision problem:

Problem: *Multiple-source CQ answering(st)*
Input: A CQ $q(\overrightarrow{x})$, sets of answers $S_1, \ldots, S_m$ to $q(\overrightarrow{x})$ w.r.t. a DL-Lite KB $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$, a tuple of individuals $\overrightarrow{a}$
Question: Does $\mathcal{K}_P \models_{\text{st}} q(\overrightarrow{a})$ hold ?

Proposition 1 *Let $\mathcal{K} = \langle \mathcal{T}, P \rangle$ be a KB with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$. $\forall st \in \{\forall, \exists, maj, \forall_c, \exists_c, maj_c\}$. Multiple-source CQ answering (st) is polynomial for data complexity.*

The proof is immediate, since by the definition of the proposed strategies, the strategy st $= \forall$ requires m calls to the *CQ answering* problem and strategies st $= \exists$ and st $= maj$, in the worst case, call it m times. Besides, computing the size of the largest sets of answers is linear in the maximum size of the ABoxes, which is bounded by a constant. Since the number of the largest sets of answers is bounded by m, computing the sets of answers of maximal cardinality is polynomial.

4 Comparative study: productivity relation between query answering strategies

Query answering strategies can be compared to each others according to different points of views: productivity, logical properties and computational complexity. In this section, we compare them from a productivity point of view; i.e., we pairwise compare the strategies in terms of answer sets inclusion, they provide w.r.t. queries. We formalize this comparison as follows:

Definition 9 Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a KB and let st_i and st_j be two strategies. We say that st_i is more productive than st_j if for any ABox profile $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$ and any query $q(\overrightarrow{x})$ the set of answers to $q(\overrightarrow{x})$, w.r.t. P, using the strategy st_j, is included or equal to the set of answers to $q(\overrightarrow{x})$, w.r.t. P, using the strategy st_i. (i.e., $S_{\mathrm{st}_j} \subseteq S_{\mathrm{st}_i}$).

In the special case of BQs, we say that a strategy st_i is more productive than a strategy st_j if for any ABox of P and any BQ $q()$ if we have $\mathcal{K}_P \models_{\mathrm{st}_j} q()$, then we also have $\mathcal{K}_P \models_{\mathrm{st}_i} q()$.

The following proposition gives an exhaustive study of productivity results for the different query answering strategies studied in this paper.

Proposition 2 *Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a KB, with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$, and $q(\overrightarrow{x})$ be a query. The comparative relation between query answering strategies $st \in \{\forall, \exists, maj, \forall_c, maj_c, \exists_c\}$, w.r.t. productivity, is depicted in Fig. 2.*

Proof – (1): $\forall s \in S_\forall$, s is such that $\forall i \in \{1, \cdots, m\}$ and $s \in S_i$, so $\frac{|\{i : \overrightarrow{s} \in S_i\}|}{m} = 1$, $s \in S_{maj}$. Namely, every answer which is derived by all ABoxes is also an answer which is derived by the majority of ABoxes.

- (2): $S_{maj} \subseteq S_\exists$. Every answer given by the majority of ABoxes is also given by one ABox. Note that by transitivity we get $S_\forall \subseteq S_\exists$.
- (3) and (4): The proofs are analogous to (1) and (2), applied to the set of answers of maximal cardinality. By transitivity, we get $S_{\forall_c} \subseteq S_{\exists_c}$.
- (5) The proof is immediate. $S_\forall \subseteq S_{\forall_c}$ and $\forall s \in S_\forall$ by definition of Sc_P, $s \in S_i$ with $S_i \in Sc_P$, so $s \in S_{\forall_c}$.
- (6) Since $S_\exists = S_1 \cup \ldots \cup S_m$, by definition of Sc_P, we have $S_{\exists_c} \subseteq S_\exists$. □

For the converse productivity relations, we provide the following counter-examples :

Example 7 (Counter-example of the converse of relations (1) and (2)) Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be the KB from Example 1. Consider the query of Example 2: $q(x) = \exists y\, TutorOf(y, x) \wedge Students(x)$.
We recall that the sets of answers to $q(x)$ from the ABoxes are :

$S_1 = \{a, b, d, e\}$
$S_2 = \{b, c, d, f\}$
$S_3 = \{b, d, e, f\}$
$S_4 = \{a, b, c\}$

Thus, we have:

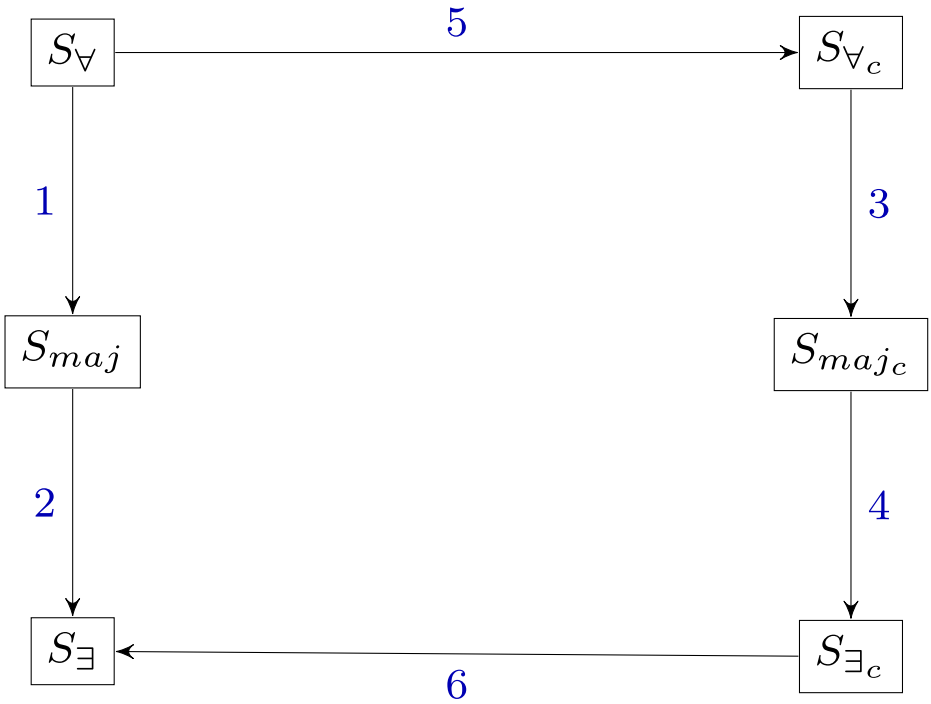


Fig. 2 $St_1 \longrightarrow St_2$ means that the set of answers obtained from the strategy St_1 is also obtained from the strategy St_2

$$\begin{array}{l} S_\forall = \{b\} \\ S_{maj} = \{b, d\} \\ S_\exists = \{a, b, c, d, e, f\} \end{array}$$

One can check that (1) $S_\forall \subseteq S_{maj}$, (2) $S_{maj} \subseteq S_\exists$, and $S_\forall \subseteq S_\exists$. However, $S_{maj} \nsubseteq S_\forall$, $S_\exists \nsubseteq S_{maj}$ and $S_\exists \nsubseteq S_\forall$.

Example 8 (Counter-example of the converse of relations (3) and (4)) Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be the KB from Example 5. Consider again the query $q(x) = \exists y\, TutorOf(y, x) \wedge Students(x)$.

We recall that the sets of answers to $q(x)$ from the ABoxes are:

$$\begin{array}{l} S_1 = \{a, b, d, e\} \\ S_2 = \{b, c, d, f\} \\ S_3 = \{b, d, e, f\} \\ S_4 = \{a, b, c\} \end{array}$$

Furthermore, $Sc_P = \{S_1, S_2, S_3\}$. As a consequence, we have:

$$\begin{array}{l} S_{\forall_c} = \{b, d\} \\ S_{maj_c} = \{b, d, e, f\} \\ S_{\exists_c} = \{a, b, c, d, e, f\} \end{array}$$

We have (3) $S_{\forall_c} \subseteq S_{maj_c}$, (4) $S_{maj_c} \subseteq S_{\exists_c}$ and $S_{\forall_c} \subseteq S_{\exists_c}$. Whereas, $S_{maj_c} \nsubseteq S_{\forall_c}$, $S_{\exists_c} \nsubseteq S_{maj_c}$ and $S_{\exists_c} \nsubseteq S_{\forall_c}$.

Example 9 (Counter-example of the converse of relations (5) and (6)) Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be the KB from Examples 2 and 5. Consider again query $q(x) = \exists y\, TutorOf(y, x) \wedge Students(x)$.
We recall that:

$$\begin{array}{l} S_\forall = \{b\} \\ S_\exists = \{a, b, c, d, e, f\} \\ S_{\forall_c} = \{b, d\} \\ S_{\exists_c} = \{a, b, c, d, e, f\} \end{array}$$

We have (5) $S_\forall \subseteq S_{\forall_c}$ and one can check that $S_{\forall_c} \nsubseteq S_\forall$.

The following counterexample confirms that $S_\exists$ is strictly more productive than the other studied inference strategies.

Example 10 ($S_\exists$ is more productive than $S_\forall, S_{maj}, S_{\forall_c}, S_{maj_c}, S_{\exists_c}$) . Let us consider the TBox $\mathcal{T}$ of Example 1:

$$\begin{array}{l} PhdStudents \sqsubseteq \neg\, Professors \\ PhdStudents \sqsubseteq Students \\ Students \sqsubseteq \exists\, TutorOf \\ TutorOf \sqsubseteq TutorialTo \end{array}$$

and the following three ABoxes:

$\mathcal{A}_1 =$ {PhdStudents(a), PhdStudents(c), TutorOf(p, a), TutorOf(p, c)},
$\mathcal{A}_2 =$ {Students(t), TutorOf(p, t)},
$\mathcal{A}_3 =$ {PhdStudents(b), Students(c), TutorOf(p, b), TutorOf(p, c)}

Let us consider the following query:
$q(x) = \exists y\, TutorOf(y, x) \wedge Students(x)$
The sets of answers to $q(x)$ from the ABoxes are:

$$\begin{array}{l} S_1 = \{a, c\} \\ S_2 = \{t\} \\ S_3 = \{b, c\} \end{array}$$

One can easily check that:
$S_\forall = \emptyset$, $S_{maj} = \{c\}$, $S_\exists = \{a, b, c, t\}$, $S_{\forall_c} = S_{maj_c} = \{c\}$, $S_{\exists_c} = \{a, b, c\}$.
Hence, we have:
$S_{\exists_c} \nsubseteq S_\forall$, $S_{\exists_c} \nsubseteq S_{maj}$, $S_{\exists_c} \nsubseteq S_{\forall_c}$, $S_{\exists_c} \nsubseteq S_{maj_c}$, $S_\exists \nsubseteq S_{\exists_c}$

The following counterexample shows that S_{maj} is incomparable with $S_{\exists_c}, S_{maj_c}$ *and* $S_{\forall_c}$.

Example 11 (S_{maj} is incomparable with $S_{\exists_c}, S_{maj_c}$ *and* $S_{\forall_c}$) Let us again consider the TBox of Example 1, namely $\mathcal{T} =$

$$\begin{array}{l} PhdStudents \sqsubseteq \neg\, Professors \\ PhdStudents \sqsubseteq Students \\ Students \sqsubseteq \exists\, TutorOf \\ TutorOf \sqsubseteq TutorialTo \end{array}$$

Let us consider the following ABoxes:

$\mathcal{A}_1 =$	{PhdStudents(a), TutorOf(p, a), TutorOf(p, c)},
$\mathcal{A}_2 =$	{Students(a), TutorOf(p, a)},
$\mathcal{A}_3 =$	{PhdStudents(b), Students(c), TutorOf(p, b), TutorOf(p, c)}

Let us consider again the query:
$q(x) = \exists y\, TutorOf(y, x) \wedge Students(x)$
The sets of answers to $q(x)$ from the ABoxes are:

$S_1 = \{a\}$
$S_2 = \{a\}$
$S_3 = \{b, c\}$

Clearly: $S_{maj} = \{a\}$ is incomparable with $S_{maj_c} = S_{\forall_c} = S_{\exists_c} = \{b, c\}$.

From a productivity point of view, universal strategies are very cautious while existential ones are very adventurous. Thereby, the majority strategy seems to be the most interesting ones.

Next section considers another criterion for comparing query answering strategies. It consists in analyzing the logical properties of each strategy.

5 Rational properties of query answering strategies

This section studies logical properties of query answering strategies. We first analyze the behavior of the introduced inferences w.r.t. the conjunction (or set union) for CQs and BQs and the consistency of the inferred conclusions for a fixed knowledge base for BQs. Besides, these inferences are inherently non-monotonic. Indeed, if any query q is entailed from a KB using a query answering strategy, then q may be questionable in the light of new factual assertions. We will assume that these new pieces of information are sure (and will refer to conditional inferences, opposed to unconditional inferences when the KB is fixed). Hence, we also analyze the proposed inferences w.r.t rationality properties introduced for non-monotonic inference that we recast in our framework.

5.1 Properties of unconditional inference

Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be KB with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$ and st $\in \{\forall, \exists, maj, \forall_c, \exists_c, maj_c\}$ be query answering strategies. We first consider the following properties introduced in [9]:

QCE (Query Conjunction Elimination) For any KB $\mathcal{K}_P$ and any CQs or BQs q_1 and q_2, if $\mathcal{K}_P \models_{\text{st}} q_1 \wedge q_2$, then $\mathcal{K}_P \models_{\text{st}} q_1$ and $\mathcal{K}_P \models_{\text{st}} q_2$.

QCI (Query Conjunction Introduction) For any KB $\mathcal{K}_P$ and any CQs or BQs q_1 and q_2, if $\mathcal{K}_P \models_{\text{st}} q_1$ and $\mathcal{K}_P \models_{\text{st}} q_2$, then $\mathcal{K}_P \models_{\text{st}} q_1 \wedge q_2$.

Cons (Consistency) For any set of assertions $\mathcal{A}'$, if $\mathcal{K}_P \models_{\text{st}} \mathcal{A}'$, then $\langle \mathcal{T}, \mathcal{A}' \rangle$ is consistent .

ConsC (Consistency of Conjunction) For any set of assertions $\mathcal{A}$, if for all $f \in \mathcal{A}$, $\mathcal{K}_P \models_{\text{st}} f$, then $\langle \mathcal{T}, \mathcal{A} \rangle$ is consistent.

Note that properties **QCE** and **QCI**, related to the conjunction property, concern CQs and BQs while the properties **Cons** and **ConsC**, related to the consistency property, only concern BQs.

Proposition 3 (Properties of unconditional inference) *The behavior of inference relations st, where st* $\in \{\forall, \exists, maj, \forall_c, \exists_c, maj_c\}$*, with respect to properties* ***QCE, QCI, Cons and ConsC****, is given in Table 1* [1].

Proof – **QCE**

Let st $= \forall$ and let q_1 and q_2 be two CQs. If $\mathcal{K}_P \models_\forall q_1(\overrightarrow{s}) \wedge q_2(\overrightarrow{s'})$, by Definition 3 $\forall S_i \in S_P, (\overrightarrow{s}, \overrightarrow{s'}) \in S_i$. Let Π_1 be the projection of the answers to query $q_1 \wedge q_2$ on the variables of q_1, and we denote by $\Pi_1(S_i)$ the set of projections of answers to query $q_1 \wedge q_2$ on the variables of q_1. Moreover, let Π_2 be the projection of the answers to query $q_1 \wedge q_2$ on the variables of q_2, and we denote by $\Pi_2(S_i)$ the set of projections of answers to query $q_1 \wedge q_2$ on the variables of q_2. $\forall \Pi_1(S_i) \in \Pi_1(S_P)$, $\overrightarrow{s} \in \Pi_1(S_i)$ and $\forall \Pi_2(S_i) \in \Pi_2(S_P)$, $\overrightarrow{s'} \in \Pi_2(S_i)$, so $\mathcal{K}_P \models_{\text{st}} q_1$ and $\mathcal{K}_P \models_{\text{st}} q_2$.
Let now q_1 and q_2 be two BQs, if $\mathcal{K}_P \models_\forall q_1() \wedge q_2()$, by Definition 7, $\forall \mathcal{A}_i \in P$, $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_1() \wedge q_2()$, then $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_1()$ and $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_2()$, so the result holds.

The proof is similar for st $\in \{\exists, maj\}$. □

Example 12 shows that st $= \forall_c$ does not satisfy **QCE**.

Example 12 (counter-example for st $= \forall_c$) Let $\mathcal{K}_P = \langle \mathcal{T}, P = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\} \rangle$ be a KB where $\mathcal{T}$ is the TBox of Example 1 and $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ are such that:

$\mathcal{A}_1$ = {PhdStudents(a), TutorOf(p, a) }
$\mathcal{A}_2$ = {PhdStudents(a), TutorOf(p, a)}
$\mathcal{A}_3$ = {PhdStudents(b), PhdStudents(c)}

Consider queries $q_1(x) = PhdStudents(x)$ and $q_2(x) = TutorOf(p, x)$. The sets of answers to $q_1(x)$ from the ABoxes are:

$S_1 = \{a\}$
$S_2 = \{a\}$
$S_3 = \{b, c\}$

Thus, $Sc_P = \{S_3\}$.
The sets of answers to $q_2(x)$ from the ABoxes are:

$S'_1 = \{a\}$
$S'_2 = \{a\}$
$S'_3 = \emptyset$

As a result, $S'c_P = \{S'_1, S'_2\}$.
The sets of answers to $q_1(x) \wedge q_2(x)$ from the ABoxes are:

$S''_1 = \{a\}$
$S''_2 = \{a\}$
$S''_3 = \emptyset$

Consequently, $S''c_P = \{S''_1, S''_2\}$.
We have $\mathcal{K}_P \models_{\forall_c} q_1(a) \wedge q_2(a)$, but $\mathcal{K}_P \not\models_{\forall_c} q_1(a)$.

Example 13 demonstrates that st $= \exists_c$ does not satisfy **QCE**.

[1] The symbols $\surd$, $\times$ (resp.) means satisfied, not satisfied (resp.)

Table 1 Properties of unconditional inference

Properties	$\forall$	$\exists$	maj	$\forall_c$	$\exists_c$	maj_c
QCE	$\surd$	$\surd$	$\surd$	$\times$	$\times$	$\times$
QCI	$\surd$	$\times$	$\times$	$\times$	$\times$	$\times$
Cons	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
ConsC	$\surd$	$\times$	$\surd$	$\surd$	$\times$	$\surd$

Example 13 (counter-example for st $= \exists_c$) Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a KB from counter-example 12. Consider again queries $q_1(x) = PhdStudents(x)$ and $q_2(x) = TutorOf(p, x)$. We have $\mathcal{K}_P \models_{\exists_c} q_1(a) \wedge q_2(a)$, but $\mathcal{K}_P \not\models_{\exists_c} q_1(a)$.

Example 14 indicates that st $= maj_c$ does not satisfy **QCE**.

Example 14 (counter-example for st $= maj_c$) Let $\mathcal{K}_P = \langle \mathcal{T}, P = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\} \rangle$ be a KB where $\mathcal{T}$ is the TBox of Example 1 and $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ are as follows:

$\mathcal{A}_1$ = {PhdStudents(a), TutorOf(p, a) }
$\mathcal{A}_2$ = {PhdStudents(a), TutorOf(p, a) }
$\mathcal{A}_3$ = {PhdStudents(b), PhdStudents(c), TutorOf(p, b)}

Consider again the queries $q_1(x) = PhdStudents(x)$ and $q_2(x) = TutorOf(p, x)$. The sets of answers to $q_1(x)$ from the ABoxes are:

$S_1 = \{a\}$
$S_2 = \{a\}$
$S_3 = \{b, c\}$

Hence, $Sc_P = \{S_3\}$.
The sets of answers to $q_2(x)$ from the ABoxes are:

$S'_1 = \{a\}$
$S'_2 = \{a\}$
$S'_3 = \{b\}$

Accordingly, $S'c_P = \{S'_1, S'_2, S'_3\}$.
The sets of answers to $q_1(x) \wedge q_2(x)$ from the ABoxes are:

$S''_1 = \{a\}$
$S''_2 = \{a\}$
$S''_3 = \{b\}$

Therefore, $S''c_P = \{S''_1, S''_2, S''_3\}$.
We have $\mathcal{K}_P \models_{maj_c} q_1(a) \wedge q_2(a)$, but $\mathcal{K}_P \not\models_{maj_c} q_1(a)$.

– **QCI**

Let st $= \forall$ and q_1 and q_2 be two CQs. If $\mathcal{K}_P \models_\forall q_1(\overrightarrow{s})$ and $\mathcal{K}_P \models_\forall q_2(\overrightarrow{s'})$, then by Definition 3, $\forall S_i \in S_P$, $\overrightarrow{s} \in S_i$ and $\forall S'_i \in S'_P$, $\overrightarrow{s'} \in S_i$, so $\forall S_i \times S'_i \in S_P \times S'_P$, $(\overrightarrow{s}, \overrightarrow{s'}) \in S_i \times S'_i$. As a consequence, $\mathcal{K}_P \models_\forall q_1(\overrightarrow{s}) \wedge q_2(\overrightarrow{s'})$. The proof is similar for st $= \forall_c$. Let now q_1 and q_2 be two BQs. If $\langle \mathcal{T}, \mathcal{A}_i \rangle \models_\forall q_1()$ and $\langle \mathcal{T}, \mathcal{A}_i \rangle \models_\forall q_2()$ by Definition 7 $\forall \mathcal{A}_i \in P$, $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_1()$ and $\forall \mathcal{A}_i \in P$, $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_2()$, then $\forall \mathcal{A}_i \in P$, $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_1() \wedge q_2()$ and the result holds.

Example 15 shows that st $= \forall_c$ does not satisfy **QCI**. [counter-example for st $= \forall_c$]

Example 15 Let $\mathcal{K}_P = \langle \mathcal{T}, P = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}\rangle$ be a KB where $\mathcal{T}$ is the TBox of Example 1 and $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ are such that:

$\mathcal{A}_1$ = {PhdStudents(a), TutorOf(p, a) }
$\mathcal{A}_2$ = {PhdStudents(a), TutorOf(p, a), TutorOf(p,b)}
$\mathcal{A}_3$ = {PhdStudents(b), PhdStudents(c)}

Consider the queries $q_1(x) = PhdStudents(x)$ and $q_2(x) = TutorOf(p, x)$. The sets of answers to $q_1(x)$ from the ABoxes are:

$S_1 = \{a\}$
$S_2 = \{a\}$
$S_3 = \{b, c\}$

Thus $Sc_P = \{S_3\}$.
The sets of answers to $q_2(x)$ from the ABoxes are:

$S'_1 = \{a\}$
$S'_2 = \{a, b\}$
$S'_3 = \emptyset$

Consequently, $S'c_P = \{S'_2\}$.
The sets of answers to $q_1(x) \wedge q_2(x)$ from the ABoxes are:

$S''_1 = \{a\}$
$S''_2 = \{a\}$
$S''_3 = \emptyset$

Thereby, $S''c_P = \{S''_1, S''_2\}$.
We have $\mathcal{K}_P \models_{\forall_c} q_1(b)$ and $\mathcal{K}_P \models_{\forall_c} q_2(b)$, but $\mathcal{K}_P \not\models_{\forall_c} q_1(b) \wedge q_2(b)$.

Example 16 shows that st $\in \{\exists, \exists_c\}$ does not satisfy **QCI**.

Example 16 (counter-example for $st \in \{\exists, \exists_c\}$) Let $\mathcal{K}_P = \langle \mathcal{T}, P = \{\mathcal{A}_1, \mathcal{A}_2\}\rangle$ be a KB where $\mathcal{T}$ is the TBox of Example 1 and $\mathcal{A}_1$ and $\mathcal{A}_2$ are such that:

$\mathcal{A}_1$ = {PhdStudents(a), TutorOf(p, a) }
$\mathcal{A}_2$ = {PhdStudents(b), TutorOf(q, b)}

Consider the queries $q_1(x) = TutorOf(p, x)$ and $q_2(y) = TutorOf(q, y)$. The sets of answers to $q_1(x)$ from the ABoxes are:

$S_1 = \{a\}$
$S_2 = \emptyset$

The sets of answers to $q_2(y)$ are:

$S'_1 = \emptyset$
$S'_2 = \{b\}$

We have $\mathcal{K}_P \models_{\exists} q_1(x)$ and $\mathcal{K}_P \models_{\exists} q_2(y)$, but the set of answers to the query $q_1(x) \wedge q_2(y)$ are:

$S''_1 = \emptyset$
$S''_2 = \emptyset$

Therefore, $\mathcal{K}_P \not\models_{\exists} q_1(x) \wedge q_2(y)$.
For BQs, consider the same knowledge base and the following BQs $q_1() = TutorOf(p, a)$

and $q_2() = TutorOf(q, b)$. We have $\mathcal{K}_P \models_\exists q_1()$ and $\mathcal{K}_P \models_\exists q_2()$, but $\mathcal{K}_P \not\models_\exists q_1() \wedge q_2()$.

Example 17 indicates that st $\in \{maj, maj_c\}$ does not satisfy **QCI**.

Example 17 (counter-example for $st \in \{maj, maj_c\}$) Let $\mathcal{K}_P = \langle \mathcal{T}, P = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}\rangle$ be a KB where T is the TBox of Example 1 and $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ are such that:

$\mathcal{A}_1$ = {PhdStudents(a), PhdStudents(c), TutorOf(p, a), TutorOf(q, c) }
$\mathcal{A}_2$ = {PhdStudents(a), PhdStudents(b), TutorOf(p, a), TutorOf(q, b),}
$\mathcal{A}_3$ = {PhdStudents(d), PhdStudents(c), TutorOf(p, d), TutorOf(q, c),}

Consider queries $q_1(x) = TutorOf(p, x)$ and $q_2(y) = TutorOf(q, y)$. The set of answers to $q_1(x)$ from the ABoxes are:

$S_1 = \{a\}$
$S_2 = \{a\}$
$S_3 = \{d\}$

The set of answers to $q_2(y)$ from the ABoxes are:

$S'_1 = \{c\}$
$S'_2 = \{b\}$
$S_3 = \{c\}$

We have $\mathcal{K}_P \models_{maj} q_1(x)$ and $\mathcal{K}_P \models_{maj} q_2(y)$, but the set of answers to $q_1(x) \wedge q_2(y)$ from the ABoxes are:

$S''_1 = \{(a, c)\}$
$S''_2 = \{(a, b)\}$
$S''_3 = \{(d, c)\}$

Hence, $\mathcal{K}_P \not\models_{maj} q_1(x) \wedge q_2(y)$.

For BQs, consider the same KB and the following BQs $q_1() = TutorOf(p, a)$ and $q_2() = TutorOf(q, c)$. We have $\mathcal{K}_P \models_{maj} q_1()$ and $\mathcal{K}_P \models_{maj} q_2()$, but $\mathcal{K}_P \not\models_{maj} q_1() \wedge q_2()$.

– **Cons**

Let st $= \forall$. Let $\mathcal{A}'$ be a set of assertions. For any $\mathcal{A}'$, if $\mathcal{K}_P \models_\forall A'$, then by Definition 8, $\forall \mathcal{A}_i \in P$, $\langle \mathcal{T}, \mathcal{A}_i\rangle \models A'$. Thus any interpretation that satisfies $\mathcal{T}$ and $\mathcal{A}_i$ satisfies A'. Consequently, $\langle \mathcal{T}, \mathcal{A}'\rangle$ is consistent.

Let st $= \exists$. Let $\mathcal{A}'$ be a set of assertions. For any $\mathcal{A}'$, if $\mathcal{K}_P \models_\exists A'$, then, by Definition 8, $\exists \mathcal{A}_i \in P$, $\langle \mathcal{T}, \mathcal{A}_i\rangle \models A'$. Thus any interpretation that satisfies $\mathcal{T}$ and $\mathcal{A}_i$ satisfies A'. As a result $\langle \mathcal{T}, \mathcal{A}'\rangle$ is consistent.

Let st $= maj$. Let $\mathcal{A}'$ be a set of assertions. For any $\mathcal{A}'$, if $\mathcal{K}_P \models_{maj} A'$, then by Definition 8, $\frac{|\{i : \langle \mathcal{T}, \mathcal{A}_i\rangle \models q()\}|}{m} > \frac{1}{2}$. Hence any interpretation that satisfies $\mathcal{T}$ and a majortity of $\mathcal{A}_i$ satisfies A'. Accordingly, $\langle \mathcal{T}, \mathcal{A}'\rangle$ is consistent.

– **ConsC**

Let st $= \forall$. For any $\mathcal{A}$, if for all $f \in \mathcal{A}$, $\mathcal{K}_P \models_\forall f$, then by Definition 8, for all $f \in \mathcal{A}$, $\forall \mathcal{A}_i \in P$, $\langle \mathcal{T}, \mathcal{A}_i\rangle \models f$. Thus for all $f \in \mathcal{A}$, $\forall \mathcal{A}_i \in P$ any interpretation that satisfies $\mathcal{T}$ and $\mathcal{A}_i$ satisfies f. Hence, $\forall \mathcal{A}_i \in P$. Any interpretation that satisfies $\mathcal{T}$ and $\mathcal{A}_i$ satisfies $\wedge f$. Therefore, $\langle \mathcal{T}, \mathcal{A}\rangle$ is consistent.

Example 18 shows that st $= \exists$ does not satisfy **ConsC**.

Example 18 (counter-example for $st = \exists$) Let $\mathcal{K}_P = \langle \mathcal{T}, P = \{\mathcal{A}_1, \mathcal{A}_2\}\rangle$ be a KB where $\mathcal{T}$ is the TBox of Example 1 and $\mathcal{A}_1$ and $\mathcal{A}_2$ are such that $\mathcal{A}_1 = \{PhdStudents(a)\}$ and $\mathcal{A}_2 = \{Professors(a)\}$. We have $\mathcal{K}_P \models_\exists PhdStudents(a)$ and $\mathcal{K}_P \models_\exists Professors(a)$. However $\langle \mathcal{T}, \{PhdStudents(a), Professors(a)\}\rangle$ is inconsistent.

Let st $= maj$. For any $\mathcal{A}$, if for all $f \in \mathcal{A}$, $\mathcal{K}_P \models_{maj} f$, then for all $f \in \mathcal{A}$, for a majority of $\mathcal{A}_i$ we have $\langle \mathcal{T}, \mathcal{A}_i\rangle \models f$. Since in DL-Lite$_R$ the conflicts are of size two by Lemma 1, we consider pairs of assertions. For any two assertions $f_1, f_2 \in \mathcal{A}$, if for a majority of $\mathcal{A}_i$ we have $\langle \mathcal{T}, \mathcal{A}_i\rangle \models f_1$ and if for a majority of $\mathcal{A}_i$ we have $\langle \mathcal{T}, \mathcal{A}_i\rangle \models f_2$, then there exists at least one A_i such that $\langle \mathcal{T}, \mathcal{A}_i\rangle \models f_1 \wedge f_2$, so $\langle \mathcal{T}, f_1, f_2\rangle$ is consistent. Therefore, $\mathcal{A}$ is pairwise consistent, and since pairwise consistency entails consistency in DL-Lite$_R$, the result holds.

5.2 Properties of conditional inference

When we deal with conflicting information, the underlying inference strategies have no standard properties. Typically, they are non-monotonic inference relations. However, even when they are monotonic, then it may happen that they fail to satisfy other properties such as the CUT property or the AND property. Therefore, it is important to study properties of these strategies independently if they are monotonic or not.

This section analyses the query answering strategies in terms of non-monotonic (or conditional/contextual) inferences. Within the propositional logic setting, several approaches have been put forward for non-monotonic (or conditional) inferences [10, 17, 18]. In such approaches, non-monotonicity is essentially caused by the fact that initial knowledge used for the inference process is incomplete.

Let $\mathcal{K}_P = \langle \mathcal{T}, P\rangle$ be a KB where $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$ and $\mathcal{A}_\alpha$ and $\mathcal{A}_\beta$ be two sets of assertions such that $\langle \mathcal{T}, \mathcal{A}_\alpha\rangle$ and $\langle \mathcal{T}, \mathcal{A}_\beta\rangle$ are consistent. Assume that $\mathcal{A}_\alpha$ is the newly added dataset that represent the context. Since $\mathcal{A}_\alpha$ is considered as more reliable than the assertions in the KB, the assertions of $\mathcal{A}_\alpha$ have to be inferred from the KB with respect to the st query answering strategies.

Now, let f and g be two formulas and $\mid\!\sim$ be a non-monotonic inference relation. Within a non-monotonic setting, $f \mid\!\sim g$ means that if f is observed or in the context of f (thus considered as valid) then g is deduced from f and from the knowledge base. In practise, within our context, if f denotes an assertion then f should be added to the ABox with highest priority. However, for hidden ABoxes, this is not possible because the ABoxes are not accessible.

In this study, we focus on the situation where the considered conclusions are sets of assertions, which can also be seen as BQs. Let q_α be a BQ such that $q_\alpha() = \wedge f$, where $f \in \mathcal{A}_\alpha$ and $q_\beta()$ are a BQ such that $q_\beta() = \wedge f$ with $f \in \mathcal{A}_\beta$.

- $\mathcal{A}_\beta$ is a universal non-monotonic consequence of $\mathcal{A}_\alpha$ w.r.t. $\mathcal{K}_P$, denoted by $\mathcal{A}_\alpha \mid\!\sim_\forall \mathcal{A}_\beta$ if
  - there exists $\mathcal{A}_i \in P$ such that $\langle \mathcal{T}, \mathcal{A}_i\rangle \models q_\alpha()$, and
  - $\forall \mathcal{A}_i \in P$, if $\langle \mathcal{T}, \mathcal{A}_i\rangle \models q_\alpha()$, then $\langle \mathcal{T}, \mathcal{A}_i\rangle \models q_\beta()$.

- $\mathcal{A}_\beta$ is an existential non-monotonic consequence of $\mathcal{A}_\alpha$ w.r.t. $\mathcal{K}_P$, denoted by $\mathcal{A}_\alpha \mid\!\sim_\exists \mathcal{A}_\beta$ if $\exists \mathcal{A}_i \in P$, $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()$, and $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\beta()$.
- $\mathcal{A}_\beta$ is a majority non-monotonic consequence of $\mathcal{A}_\alpha$ w.r.t. $\mathcal{K}_P$, denoted by $\mathcal{A}_\alpha \mid\!\sim_{maj} \mathcal{A}_\beta$ if
  - there exists $\mathcal{A}_i \in P$ such that $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()$, and
  - $\frac{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \text{ and } \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\beta()\}|}{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()\}|} > \frac{1}{2}$.

One can check that the majority relation $\mid\!\sim_{maj}$ is non-monotonic. Indeed, let us consider the following simple counter-example where $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, $\mathcal{T} = \{B \sqsubseteq \neg C\}$, $\mathcal{A}_1 = \{B(e), D(e)\}$, $\mathcal{A}_2 = \{C(e), D(e)\}$, $\mathcal{A}_3 = \{B(e)\}$. Clearly $B(e)$ is a majority consequence of $\mathcal{T}$ (since two ABoxes over 3 support $B(e)$), while $B(e)$ is not a consequence of $\mathcal{T}$ augmented by $\{D(e)\}$ (since over the two ABoxes that support $D(e)$, only one of them supports $B(e)$). The two remaining inferences relations $\mid\!\sim_\forall$ and $\mid\!\sim_\exists$ are non-monotonic neither due the fact that the definitions require the existence of an Abox $\mathcal{A}_i$ such that $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()$ when defining $\mathcal{A}_\alpha \mid\!\sim_\forall \mathcal{A}_\beta$ ($\mathcal{A}_\alpha \mid\!\sim_\exists \mathcal{A}_\beta$). Therefore if $\mathcal{A}_\alpha \mid\!\sim_\forall \mathcal{A}_\beta$ then this does not necessarily imply that $\mathcal{A}_\alpha \cup \mathcal{A}_\gamma \mid\!\sim_\forall \mathcal{A}_\beta$, if $\mathcal{A}_\alpha \cup \mathcal{A}_\gamma$ contradicts the TBox $\mathcal{T}$. For instance, assume that $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ with $\mathcal{T} = \{A \sqsubseteq \neg B, A \sqsubseteq C\}$, $\mathcal{A}_1 = \{A(e)\}$. Let $\mathcal{A}_\alpha = \{A(e)\}$, $\mathcal{A}_\beta = \{C(e)\}$. Once can check that $\mathcal{A}_\alpha \mid\!\sim_\forall \mathcal{A}_\beta$ but $\mathcal{A}_\alpha \cup \mathcal{A}_\gamma \not\mid\!\sim_\forall \mathcal{A}_\beta$ for $\mathcal{A}_\gamma = \{B(e)\}$, since $\langle \mathcal{T}, \mathcal{A}_1 \rangle \models B(e) \cap C(e)$ does not hold. Note that if we drop condition (i) (namely the existence of $\mathcal{A}_i$ such that $\langle \mathcal{T}, \mathcal{A}_i \rangle \models \alpha$) the inferences relations $\mid\!\sim_{maj}$ and $\mid\!\sim_\forall$ become either trivial or undefined. Besides, $\mid\!\sim_\forall$ is monotonic in this case.
Let us now study the properties of our inference relations.

We first rephrase within our framework some Kraus, Lehmann and Maqidor (KLM) rationality properties [10].

Let $\mathcal{A}_\alpha$, $\mathcal{A}_\beta$ and $\mathcal{A}_\gamma$ be consistent sets of assertions, w.r.t. $\mathcal{T}$, and $\mid\!\sim$ be an inference relation. The *KLM* logical properties that we consider are the following:

R (Reflexivity) If $\exists \mathcal{A}_i \in P$ such that $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()$, then $\mathcal{A}_\alpha \mid\!\sim \mathcal{A}_\alpha$.
LLE (Left Logical Equivalence) If $\langle \mathcal{T}, \mathcal{A}_\alpha \rangle \equiv \langle \mathcal{T}, \mathcal{A}_\beta \rangle$ and $\mathcal{A}_\alpha \mid\!\sim \mathcal{A}_\gamma$, then $\mathcal{A}_\beta \mid\!\sim \mathcal{A}_\gamma$.
RW (Right Weakening) If $\langle \mathcal{T}, \mathcal{A}_\alpha \rangle \models \langle \mathcal{T}, \mathcal{A}_\beta \rangle$ and $\mathcal{A}_\gamma \mid\!\sim \mathcal{A}_\alpha$, then $\mathcal{A}_\gamma \mid\!\sim \mathcal{A}_\beta$.
Cut If $\mathcal{A}_\alpha \mid\!\sim \mathcal{A}_\beta$ and $\mathcal{A}_\alpha \cup \mathcal{A}_\beta \mid\!\sim \mathcal{A}_\gamma$, then $\mathcal{A}_\alpha \mid\!\sim \mathcal{A}_\gamma$.
CM (Cautious Monotony) If $\mathcal{A}_\alpha \mid\!\sim \mathcal{A}_\beta$ and $\mathcal{A}_\alpha \mid\!\sim \mathcal{A}_\gamma$, then $\mathcal{A}_\alpha \cup \mathcal{A}_\beta \mid\!\sim \mathcal{A}_\gamma$.
And If $\mathcal{A}_\alpha \mid\!\sim \mathcal{A}_\beta$ and $\mathcal{A}_\alpha \mid\!\sim \mathcal{A}_\gamma$, then $\mathcal{A}_\alpha \mid\!\sim \mathcal{A}_\beta \cup \mathcal{A}_\gamma$.

R means that if an assertion is a consequence of at least one ABox, it has to be a consequence of the inference relation. **LLE** expresses the fact that two equivalent sets of assertions induce the same set of consequences. **RW** says that the consequences of plausible assertions are plausible assertions too. **Cut** indicates that if a plausible consequence is as sure as the assumptions it is based on, then it can be added into the assumptions. **CM** shows that learning new assertions that can be plausibly inferred should not invalidate previous consequences. **And** highlights that the conjunction of two plausible consequences is a plausible consequence. The next proposition summarizes properties satisfied by our inference relations.

Proposition 4 (Properties of conditional inference) *The behavior of inference relations* $\mid\!\sim_{st}$ *with* $st \in \{\forall, \exists, maj\}$ *with respect to properties* ***R, LLE, RW, Cut, CM, And*** *is given in Table 2*[2].

[2]The symbols $\surd$, $\times$ (resp.) means satisfied, not satisfied (resp.)

Table 2 Properties of conditional inference $\{\forall, \exists, maj\}$

Properties	$\mid\!\sim_{\forall}$	$\mid\!\sim_{\exists}$	$\mid\!\sim_{maj}$
R	$\surd$	$\surd$	$\surd$
LLE	$\surd$	$\surd$	$\surd$
RW	$\surd$	$\surd$	$\surd$
Cut	$\surd$	$\surd$	$\times$
CM	$\surd$	$\times$	$\times$
And	$\surd$	$\times$	$\times$

Proof – **R** directly follows from the definition of non-monotonic inferences.

– **LLE**
st $= \forall$. If $\mathcal{A}_\alpha \mid\!\sim_{\forall} \mathcal{A}_\gamma$, then $\forall \mathcal{A}_i \in P$. If $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()$, then $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\gamma()$. Thus, $\forall \mathcal{A}_i \in P$, for any interpretation that satisfies $\mathcal{T}$ and $\mathcal{A}_i$, and $q_\alpha()$ also satisfies $q_\gamma()$. Since by hypothesis $\langle \mathcal{T}, \mathcal{A}_\alpha \rangle \equiv \langle \mathcal{T}, \mathcal{A}_\beta \rangle$ and as $q_\alpha() = \wedge f$ with $f \in \mathcal{A}_\alpha$ and $q_\beta() = \wedge f$ with $f \in \mathcal{A}_\beta$, we have $\forall \mathcal{A}_i \in P$, for any interpretation that satisfies T, $\mathcal{A}_i$ and $q_\beta()$ as well as $q_\gamma()$. Therefore, $\mathcal{A}_\beta \mid\!\sim_{\forall} \mathcal{A}_\gamma$. The proof is similar for st $\in \{\exists, maj\}$.

– **RW**
st $= \forall$. If $\mathcal{A}_\gamma \mid\!\sim_{\forall} \mathcal{A}_\alpha$, then $\forall \mathcal{A}_i \in P$. If $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\gamma()$, then $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()$. Hence, $\forall \mathcal{A}_i \in P$, for any interpretation that satisfies $\mathcal{T}$, $\mathcal{A}_i$ and $q_\gamma()$ as well as $q_\alpha()$. Since by hypothesis $\langle \mathcal{T}, \mathcal{A}_\alpha \rangle \models \langle \mathcal{T}, \mathcal{A}_\beta \rangle$ and because $q_\alpha() = \wedge f$ where $f \in \mathcal{A}_\alpha$ and $q_\beta() = \wedge f$ with $f \in \mathcal{A}_\beta$, we have $\forall \mathcal{A}_i \in P$, for any interpretation that satisfies T, $\mathcal{A}_i$ and $q_\gamma()$ as well as $q_\beta()$. Thereby, $\mathcal{A}_\gamma \mid\!\sim_{\forall} \mathcal{A}_\alpha$. The proof is similar for st $\in \{\exists, maj\}$.

– **Cut**
st $= \forall$. If $\mathcal{A}_\alpha \mid\!\sim_{\forall} \mathcal{A}_\beta$ and $\mathcal{A}_\alpha \cup \mathcal{A}_\beta \mid\!\sim_{\forall} \mathcal{A}_\gamma$, then $\forall \mathcal{A}_i \in P$. If $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()$, then $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\beta()$. If $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \cup q_\beta()$, then $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\gamma()$. Thus, any interpretation that satisfies $\mathcal{T}$, $\mathcal{A}_i$ and $q_\alpha()$ also satisfies $q_\beta()$. If any interpretation that satisfies $q_\alpha()$ and $q_\alpha()$ satisfies $q_\gamma()$, then any interpretation that satisfies $q_\alpha()$ also satisfies $q_\gamma()$, so $\mathcal{A}_\alpha \mid\!\sim_{\forall} \mathcal{A}_\gamma$.
st $= \exists$. If $\mathcal{A}_\alpha \mid\!\sim_{\exists} \mathcal{A}_\beta$ and $\mathcal{A}_\alpha \cup \mathcal{A}_\beta \mid\!\sim_{\exists} \mathcal{A}_\gamma$, then $\exists \mathcal{A}_i \in P$. If $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()$, then $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\beta()$. Thus, $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \cup q_\beta()$ by the second hypothesis $\exists \mathcal{A}_i \in P$, such that$\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\gamma()$. As a result, $\exists \mathcal{A}_i \in P$. If $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()$, then $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\gamma()$, so $\mathcal{A}_\alpha \mid\!\sim_{\exists} \mathcal{A}_\gamma$. □

We now give a counterexample for the majority strategy:

Example 19 Let $\mathcal{K}_P = \langle \mathcal{T}, P = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4\} \rangle$ be a KB where T is the TBox of Example 1 and $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_4$ are such that:

$$\begin{aligned}
\mathcal{A}_1 = \; & \{PhdStudents(a), PhdStudents(f), PhdStudents(e), PhdStudents(g), \\
& TutorOf(p, a), TutorOf(q, f), TutorOf(q, e), TutorOf(q, g)\} \\
\mathcal{A}_2 = \; & \{PhdStudents(b), PhdStudents(f), PhdStudents(e), PhdStudents(g), \\
& TutorOf(p, b), TutorOf(q, f), TutorOf(q, e), TutorOf(q, g)\} \\
\mathcal{A}_3 = \; & \{PhdStudents(c), PhdStudents(f), PhdStudents(e), TutorOf(p, c), \\
& TutorOf(q, f), TutorOf(q, e)\} \\
\mathcal{A}_4 = \; & \{PhdStudents(d), PhdStudents(f), TutorOf(p, d), TutorOf(q, f)\}
\end{aligned}$$

Connsider BQs $q_\alpha() = TutorOf(q, f)$, $q_\beta() = TutorOf(q, e)$ and $q_\gamma() = TutorOf(q, g)$. We have

$\mathcal{A}_\alpha \mid\!\sim_{maj} \mathcal{A}_\beta$ since
$\frac{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \text{ and } \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\beta()\}|}{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()\}|} = \frac{3}{4}$ and
$\mathcal{A}_\alpha \cup \mathcal{A}_\beta \mid\!\sim_{maj} \mathcal{A}_\gamma$ because
$\frac{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \cup q_\beta() \text{ and } \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\gamma()\}|}{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \cup q_\beta()\}|} = \frac{2}{3}$. However, $\mathcal{A}_\alpha \not\mid\!\sim_{maj} \mathcal{A}_\gamma$ as
$\frac{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \text{ and } \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\gamma()\}|}{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()\}|} = \frac{1}{2}$.

– **CM**

st $= \forall$. If $\mathcal{A}_\alpha \mid\!\sim_\forall \mathcal{A}_\beta$ and $\mathcal{A}_\alpha \mid\!\sim_\forall \mathcal{A}_\gamma$, then $\forall \mathcal{A}_i \in P$, if $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()$, then $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\beta()$ and $\forall \mathcal{A}_i \in P$, if $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()$ then $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\gamma()$. Thus $\forall \mathcal{A}_i \in P$, any interpretation that satisfies $\mathcal{T}$, $\mathcal{A}_i$ and $q_\alpha()$ also satisfies $q_\beta()$ and $q_\gamma()$. Hence, $\forall \mathcal{A}_i \in P$, any interpretation that satisfies $\mathcal{T}$, $\mathcal{A}_i$ and $q_\alpha() \cup q_\beta()$ also satisfies $q_\gamma()$. Therefore, $\mathcal{A}_\alpha \cup \mathcal{A}_\beta \mid\!\sim_\forall \mathcal{A}_\gamma$. We now give a counterexample when the existantial strategy fails to satisfy CM:

Example 20 Let $\mathcal{K}_P = \langle \mathcal{T}, P = \{\mathcal{A}_1, \mathcal{A}_2\} \rangle$ be a KB where T is the TBox of Example 1 and and $\mathcal{A}_1$ and $\mathcal{A}_2$ are as follows:

$\mathcal{A}_1 =$	$\{PhdStudents(a), PhdStudents(b), PhdStudents(d), TutorOf(p, a), TutorOf(p, b), TutorOf(q, d)\}$
$\mathcal{A}_2 =$	$\{PhdStudents(b), PhdStudents(c), PhdStudents(e), TutorOf(p, b), TutorOf(q, c), TutorOf(q, e)\}$

Consider BQs $q_\alpha() = TutorOf(p, b)$, $q_\beta() = TutorOf(q, d)$ and $q_\gamma() = TutorOf(q, e)$.
We have $\mathcal{A}_\alpha \mid\!\sim_\exists \mathcal{A}_\beta$ since if $\langle \mathcal{T}, \mathcal{A}_1 \rangle \models q_\alpha()$, then $\langle \mathcal{T}, \mathcal{A}_1 \rangle \models q_\beta()$, we also have $\mathcal{A}_\alpha \mid\!\sim_\exists \mathcal{A}_\gamma$ because if $\langle \mathcal{T}, \mathcal{A}_2 \rangle \models q_\alpha()$, then $\langle \mathcal{T}, \mathcal{A}_2 \rangle \models q_\gamma()$. However
$\mathcal{A}_\alpha \cup \mathcal{A}_\beta \not\mid\!\sim_\exists \mathcal{A}_\gamma$ as $\langle \mathcal{T}, \mathcal{A}_2 \rangle \not\models q_\alpha() \cup q_\beta()$ and $\langle \mathcal{T}, \mathcal{A}_1 \rangle \models q_\alpha() \cup q_\beta()$, but $\langle \mathcal{T}, \mathcal{A}_1 \rangle \not\models q_\gamma()$.

We now give a counterexample when the majority strategy fails to satisfy CM:

Example 21 Let $\mathcal{K}_P = \langle \mathcal{T}, P = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\} \rangle$ be a KB where T is the TBox of Example 1 and $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ are as follows:

$\mathcal{A}_1 =$	$\{PhdStudents(a), PhdStudents(d), PhdStudents(e), TutorOf(p, a), TutorOf(q, d), TutorOf(q, e)\}$
$\mathcal{A}_2 =$	$\{PhdStudents(b), PhdStudents(d), PhdStudents(e), PhdStudents(f), TutorOf(p, b), TutorOf(q, d), TutorOf(q, e), TutorOf(q, f)\}$
$\mathcal{A}_3 =$	$\{PhdStudents(c), PhdStudents(d), PhdStudents(f), TutorOf(p, c), TutorOf(q, d), TutorOf(q, f)\}$

Consider the BQs $q_\alpha() = TutorOf(q, d)$, $q_\beta() = TutorOf(q, e)$ and $q_\gamma() = TutorOf(q, f)$. We have
$\mathcal{A}_\alpha \mid\!\sim_{maj} \mathcal{A}_\beta$ since
$\frac{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \text{ and } \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\beta()\}|}{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()\}|} = \frac{2}{3}$ and
$\mathcal{A}_\alpha \mid\!\sim_{maj} \mathcal{A}_\gamma$ because
$\frac{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \text{ and } \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\gamma()\}|}{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()\}|} = \frac{2}{3}$.
However, $\mathcal{A}_\alpha \cup \mathcal{A}_\beta \not\mid\!\sim_{maj} \mathcal{A}_\gamma$ as
$\frac{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \cup q_\beta() \text{ and } \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\gamma()\}|}{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \cup q_\beta()\}|} = \frac{1}{2}$.

– **And**

st $= \forall$. If $\mathcal{A}_\alpha \mid\!\sim_\forall \mathcal{A}_\beta$ and $\mathcal{A}_\alpha \mid\!\sim_\forall \mathcal{A}_\gamma$ then $\forall \mathcal{A}_i \in P$, if $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()$ then $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\beta()$, and $\forall \mathcal{A}_i \in P$, if $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()$ then $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\gamma()$. Thus, $\forall \mathcal{A}_i \in P$, any interpretation that satisfies $\mathcal{T}$, $\mathcal{A}_i$ and $q_\alpha()$ also satisfies $q_\beta()$ and $q_\gamma()$, as a consequence $\forall \mathcal{A}_i \in P$, any interpretation that satisfies $\mathcal{T}$, $\mathcal{A}_i$ and $q_\alpha()$ also satisfies $q_\beta() \cup q_\gamma()$. Therefore $\mathcal{A}_\alpha \mid\!\sim_\forall \mathcal{A}_\beta \cup \mathcal{A}_\gamma$. We now give a counterexample for the existantial strategy:

Example 22 Let $\mathcal{K}_P = \langle \mathcal{T}, P = \{\mathcal{A}_1, \mathcal{A}_2\} \rangle$ be the KB of Example 20. We have $\mathcal{A}_\alpha \mid\!\sim_\exists \mathcal{A}_\beta$ since if
$\langle \mathcal{T}, \mathcal{A}_1 \rangle \models q_\alpha()$, then $\langle \mathcal{T}, \mathcal{A}_1 \rangle \models q_\beta()$, and we have $\mathcal{A}_\alpha \mid\!\sim_\exists \mathcal{A}_\gamma$ because if $\langle \mathcal{T}, \mathcal{A}_2 \rangle \models q_\alpha()$ then $\langle \mathcal{T}, \mathcal{A}_2 \rangle \models q_\gamma()$. Whereas, $\mathcal{A}_\alpha \not\mid\!\sim_\exists \mathcal{A}_\beta \cup \mathcal{A}_\gamma$ as $\langle \mathcal{T}, \mathcal{A}_1 \rangle \not\models q_\beta() \cup q_\gamma()$ and $\langle \mathcal{T}, \mathcal{A}_2 \rangle \not\models q_\beta() \cup q_\gamma()$.

We now give a counterexample for the majority strategy:

Example 23 Let $\mathcal{K}_P = \langle \mathcal{T}, P = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\} \rangle$ be the KB of Example 21. We have $\mathcal{A}_\alpha \mid\!\sim_{maj} \mathcal{A}_\beta$ since $\frac{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \text{ and } \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\beta()\}|}{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()\}|} = \frac{2}{3}$ and we have $\mathcal{A}_\alpha \mid\!\sim_{maj} \mathcal{A}_\gamma$ because $\frac{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \text{ and } \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\gamma()\}|}{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()\}|} = \frac{2}{3}$. Nevertheless, $\mathcal{A}_\alpha \not\mid\!\sim_{maj} \mathcal{A}_\beta \cup \mathcal{A}_\gamma$ as $\frac{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha() \text{ and } \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\beta() \cup q_\gamma()|}{|\{i : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q_\alpha()\}|} = \frac{1}{3}$.

From logical-property standpoint, universal strategies satisfy all the properties of unconditional and conditional inferences, which is not the case for existential and majority ones. The majority strategies satisfy more properties of unconditional inferences than existential ones.

6 Related work

There have been several research lines related to our work. The first line is belief merging, which has been largely studied in the setting of propositional logics [19–24] and more recently in the setting of description logics [5, 6, 25, 26]. Such work aims to merge knowledge bases with the purpose of getting a global view that requires knowing the whole content of the knowledge base before performing merging. Merging pieces of information need to use some fusion operators that permit combining them while respecting different constraints between sources. Our paper differs from these studies since, the problem of query answering is done without merging datasources nor knowing their contents.

The second line concerns the framework of databases querying sources, which presents some limitations investigated using views [27–29]. However these approaches mainly concern query optimization or rewriting and query languages.

The third research line related to our work is inconsistency-tolerant query answering, which is handled in the framework of database approaches [30–34], propositional logic approaches [17, 35–37] or description logics [38], and recently in lightweight description logic DL-Lite [39–42]. Such works have the objective of solving the problem of inconsistency by mainly computing a set of consistent subsets of assertional facts called repairs, which recover consistency with respect to the ontology, hence using them to carry out query answering.

The fourth research line is querying web data bases without accessing to data [15, 43, 44]. Such work has proposed methods and heuristics to efficiently query distributed ontological data based on a series of properties of summarized data.

The fifth research line which is related to our work concerns the growing interest in techniques for hiding parts of the signature of an ontology [45–47]. Such work has put forward a new algorithm that makes it easy to ask deep Web databases. Our work differs form that, since it has been assumed that each ABox in our knowledge base is consistent w.r.t. the TBox and the responses to queries as well.

7 Discussions

The different strategies defined above (in particular the existential and universal strategies) have been used in some reasoning frameworks, such as for example in default reasoning or in the framework of reasoning in the presence of inconsistencies. However, the context in which these strategies are used in this paper is different. In existing work, it is assumed that the set of so-called default extensions or repairs (defined as maximum coherent subsets of an initial potentially inconsistent knowledge base) are all mutually exclusive. In this context, the ∃-strategy relation is considered as adventurous because it systematically leads to an inconsistent set of conclusions. However, the ∀-strategy relation is considered prudent and safe because it only considers valid conclusions that have the endorsement of each extension. In our context, (hidden) ABoxes are not necessarily mutually exclusive. As a result, the ∀-strategy relation remains sound and prudent. However, the ∃-strategy relation should not be systematically discarded because if the union of the ABoxes is consistent with the TBox then the ∃-strategy relation fully makes sense and may be even preferred to ∀-strategy relation. Indeed, assume that the TBox $\mathcal{T} = \emptyset$ is empty and one have three ABoxes $\mathcal{A}_1 = \{R(a, c)\}, \mathcal{A}_2 = \{R(a, c)\}, \mathcal{A}_3 = \{R(d, a)\}$. Assume that we have the following query $q(x, y) = R(x, y)$.

We have $S_\forall = \emptyset, S_{maj} = \{(a, c)\}, S_\exists = \{(a, c), (d, a)\}$. In this situation, since $\{R(a, c), R(d, a)\}$ is consistent with $\mathcal{T} = \emptyset$, it may sense to privilege ∃-strategy relation. The situation would be different if the TBox is equal to $\mathcal{T} = \{\exists R \sqsubseteq \neg\exists R^-\}$ (namely, a negative axiom that excludes individual to be in both columns of a relation). In this situation, the use of ∃-strategy relation is questionable since $\{R(a, c), R(d, a)\}$ is inconsistent with $\mathcal{T}$. The concern is that the information on the consistency of the union of these hidden ABoxes, with respect to the TBox, is unknown. The only information which can guide us is the test of the coherence of the set of quey answers with respect to the TBox. If this set is inconsistent then clearly ∃-strategy relation cannot be safely used. If it is not the case, then the ∃-strategy relation can be used to get a larger set of responses (than the one obtained by ∀-strategy relation). Besides, from studied rational properties, the inference relations studied in this paper have different behaviors. The universal inference strategy is safe and satisfies all rational properties. As it is said above, it is safe since it only derives unquestionable answers even if the union of ABoxes is inconsistent with the TBox. However, ∀-strategy relation is cautious. The two principles proposed in the paper, cardinality selection and/or majority strategies, allow to derive more conclusions while still guaranteeing the consistency of the answer sets. However, the use of the majority principle has made it possible to obtain the satisfaction of more properties of rationality than if the principle of cardinality was used (see Table 1). Lastly, the existential strategies should be done in a careful way.

In summary, on the basis of above discussions, to one may associate degrees of confidence (or a degree that expresses to what extent a conclusion is safe) $1 > a > b > c > d > e > f > 0$ with the conclusions obtained by the different strategies in the following way: i) degree 1 is associated with the conclusions obtained with ∀-strategy relation will have degree 1, ii) degree "a" is associated with conclusions obtained from S_{maj}, iii) degree "b" is associated with conclusions obtained with $S_{\forall_c}$, iv) degree "c" is associated with conclusions obtained with S_{maj_c}, v) the degree "d" is associated with the conclusions obtained with $S_{\exists_c}$ if $S_{\exists_c}$ is consistent with the TBox, vi) the degree "e" is associated with the conclusions obtained with $S_{\exists}$ if the query answsers obtained using $S_{\exists}$ are consistent with the TBox, and vii) the degree "f" is associated with the conclusions obtained with $S_{\exists}$ or $S_{\exists_c}$ if the query answsers obtained using $S_{\exists_c}$ are inconsistent with the TBox.

Let us now briefly discuss how to go beyond the strategies defined in this paper.

The first idea is to use a richer language than DL-Lite. From semantic point of view, there is no difficulty to adapt our strategies. Indeed, the expressiveness of the language more affects the TBox rather than the Abox. However, from computational point of view, one may lost the tractability of DL-Lite. This may first has an impact on the computational complexity of checking the consistency of the result with respect to the TBox.

Indeed, as explained above one way to choose between the strategies (for instance between Maj-strategy relation, ∃-strategy relation and ∀-strategy relation) is to check the consistency of the query results with respect to the TBox. Again, a said above, in our context, the use of ∃-strategy relation may make sense only if query results are consistent with the TBox (since the union hidden ABoxes may be consistent).

Therefore, the use of ∃-strategy relation needs first to check that query results are consistent with the TBox. Such consistency check may be untractable in general description logics. Lastly, when using general DL, the set of conflicts are no longer binary. In this case, it is known that the consistency of S_{maj} result is no longer guaranteed. For instance, assume that we use a language where the negative axiom $A \sqcap B \sqcap C \sqsubseteq \bot$ is accepted (namely, there is no individual that belongs to three concepts A, B and C). Assume that $\mathcal{A}_1 = \{A(a), A(b)\}, \mathcal{A}_2 = \{A(a), A(c)\}, \mathcal{A}_3 = \{A(b), A(c)\}, \mathcal{T} = A \sqcap B \sqcap C \sqsubseteq \bot$. Assume that $q(x) = A(x)$. Clearly, $A(a)$, $A(b)$ and $A(c)$ are all majority consequences of $\langle \mathcal{T}, P = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}\rangle$. However, $\{A(a), A(b), A(c)\}$ is inconsistent.

Another way to go beyond our strategies is to define new criteria to descriminate between query answers sets. The idea is extend the cardinality criterion used in our strategies to take into account weights associated with each possible answer. The weight may represent the number of "supports" associated with each query answer. This may be obtained by rewriting a query, by replacing "∃" in the query by "∀". For instance, assume that we have: $\mathcal{T} = \{A \sqsubseteq B, \exists R \sqsubseteq \neg\exists R^-\}, \mathcal{A}_1 = \{A(a), R(b, a), R(c, a)\}, \mathcal{A}_2 = \{A(e), R(a, e)\}$.

Assume that we have the following query: $q(x) = \exists y R(y, x) \wedge B(x)$. We then have $S_1 = \{(a)\}$, $S_2 = \{(e)\}$ the two query answer sets have the same cardinality. However, the answer "$q(a)$" has two "supports" ($\{A(a), R(b, a)\}$, $\{A(a), R(c, a)\}$) while the answer "$q(e)$" has only one "support" $\{A(a), R(a, e)\}$.

In this case, one may privilege S_1 (or $q(a)$) as a definite answer of $q(x) = \exists y R(y, x) \wedge B(x)$.

One way to count the number of supports is to reformulate a query by removing the quantifier ∃ (that bears on bounded variables).

More precisely, let $q(\overrightarrow{x}) = \forall \overrightarrow{x}.\exists \overrightarrow{y}\, C(\overrightarrow{x}, \overrightarrow{y})$ a conjunctive query. Let $q(\overrightarrow{e})$ be an answer of $q(\overrightarrow{x})$ using the TBox $\mathcal{T}$ and the Abox $\mathcal{A}_i$. Let $q(\overrightarrow{e})$ be an answer of $q(\overrightarrow{x})$.

Then define $\alpha_i(\overrightarrow{e})$ as the number of supports of $q(\overrightarrow{e})$ as the cardinality of the answer set associated with the query $q(\overrightarrow{y}) = \forall \overrightarrow{y}\, C(\overrightarrow{e}, \overrightarrow{y})$ from the TBox $\mathcal{T}$ and the Abox $\mathcal{A}_i$. The answer sets S_i now contains weighted answers $(q(\overrightarrow{e}), \alpha_i(\overrightarrow{e}))$.

On the basis of these numbers of supports, one can define a preference relation between answer sets as follow : $S_i >_s S_j$ iff $\sum\{\alpha_i(\overrightarrow{e})\ (q_i(\overrightarrow{e}), \alpha_i(\overrightarrow{e})) \in S_i\} > \sum\{\alpha_j(\overrightarrow{e})\ (q_j(\overrightarrow{e}), \alpha_j(\overrightarrow{e})) \in S_j\}$. This allows to define a relation over S_p, denoted by $S_{support}$, as follows:

$$S_{support} = \{S_i : \exists S_j \in S_p, S_j >_s S_i\}$$

Once $S_{support}$ is computed, the inferences strategies $\forall, \exists, maj$ can be defined in a same spirit as the ones defined using cardinality criterion.

8 Conclusion

In this paper, we investigated the query answering problem when the content of the ABox (Data) is unknown. We proposed several inference strategies from the sets of answers provided by the ABoxes. We considered the cardinality of the sets of answers in defining new cardinality-based inferences. We compared these inferences in terms of productivity. We then studied how the propounded strategies behave with respect to the logical properties concerning the conjunction and consistency of queries and with respect to KLM rationality properties rephrased within our framework. We lastly provided a computational complexity analysis that shows that these inferences are achieved in polynomial time.

Productivity, logical properties and computational complexity provide three points of view to compare the inferences. The computational complexity standpoint does not help to compare universal, existential and majority strategies. Productivity indicates that universal strategies are very cautious. Yet, they satisfy all the properties of unconditional and conditional inferences. In constrast, existential strategies are more productive than the universal ones, but they do not satisfy all logical properties. Productivity clearly shows that majority strategies can be considered a good compromise between these two extreme strategies. This is confirmed by the logical standpoint since they satisfy more properties of unconditional inferences than existential ones.

Acknowledgments This work has received support from the European Project H2020 Marie Sklodowska-Curie Actions (MSCA), Research and Innovation Staff Exchange (RISE): Aniage project (High Dimensional Heterogeneous Data based Animation Techniques for Southeast Asian Intangible Cultural Heritage Digital Content), project number 691215.
The third author also received support from the AAP A2U QUID (Querying heterogeneous Data) project.

References

1. Baader, F., Calvanese, D., McGuinness, D., Daniele, N., Nardi, D., Patel-Schneider, P.F.: The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, New York (2003)
2. Lenzerini, M.: Ontology-Based data management. In: Proceedings of the 6th Alberto Mendelzon International Workshop on Foundations of Data Management, vol. 866, pp. 12–15. ACM, Glasgow (2011)
3. Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M., Rosati, R.: Linking data to ontologies. J. Data Semantics **4900**, 133–173 (2008)

4. Rodriguez-Muro, M., Kontchakov, R., Zakharyaschev, M.: Ontology-based data access: Ontop of databases. In: 12Th International Semantic Web Conference, Volume 8218, pp. 558–573. Springer, Sydney (2013)
5. Benferhat, S., Bouraoui, Z., Lagrue, S., Rossit, J.: Min-based assertional merging approach for prioritized DL-lite knowledge bases. In: 8Th International Scalable Uncertainty Management Conference, vol. 8720, pp. 8–21. Springer, Oxford (2014)
6. Wang, Z., Wang, K., Jin, Y., Qi, G.: Ontomerge a system for merging DL-lite ontologies. In: CEUR Workshop Proceedings, vol. 969, pp. 16–27 (2014)
7. Benferhat, S., Bouraoui, Z., Chau, M., Lagrue, S., Rossit, J.: A polynomial algorithm for merging lightweight ontologies in possibility theory under incommensurability assumption. In: Proceedings of the 9th International Conference on Agents and Artificial Intelligence, vol. 2, pp. 415–422. SciTe Press, Porto (2017)
8. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable Reasoning and Efficient Query Answering in Description logics: The DL-lite Family. J. Autom. Reason. **39**, 385–429 (2007)
9. Baget, J.F., Benferhat, S., Bouraoui, Z., Croitoru, M., Mugnier, M., Papini, O., Rocher, S., Tabia, K.: Inconsistency-tolerant query answering: rationality properties and computational complexity analysis. In: 15Th European Conference of Logics in Artificial Intelligence, vol. 10021, pp. 64–80. Larnaca, Cyprus (2016)
10. Kraus, S., Lehmann, D.J., Magidor, M.: Nonmonotonic Reasoning, Preferential Models and Cumulative Logics, CoRR - Computing Research Repository, Volume cs.AI/0202021 (2002)
11. Artale, A., Calvanese, D., Kontchakov, R., Zakharyaschev, M.: The DL-lite Family and Relations, coRR - Computing Research Repository, Volume arXiv:abs/1401.3487 (2014)
12. Flouris, G., Huang, Z., Pan, J.Z., Plexousakis, D., Wache, H.: Inconsistencies, negations and changes in ontologies. In: Proceedings, the Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, pp. 1295–1300. AAAI Press, Boston (2006)
13. Nonfjall, H., Larsen, H.L.: Detection of potential inconsistencies in knowledge bases. Int. J. Intel. Syst. **7**, 81–96 (1992)
14. Calvanese, D., Kharlamov, E., Nutt, W., Zheleznyakov, D.: Evolution of DL-lite knowledge bases. In: 9Th International Semantic Web Conference, pp. 112–128. Springer, Shanghai (2010)
15. Hamdi, G., Omri, M.N., Papini, O., Benferhat, S., Bouraoui, Z.: Querying DL-lite knowledge bases from hidden datasets. In: International Symposium on Artificial Intelligence and Mathematics, Fort Lauderdale, Florida, USA (2018)
16. Reiter, R.: A theory of diagnosis from first principles. Artificial Intelligence Journal **32**, 57–95 (1987)
17. Benferhat, S., Cayrol, C., Dubois, D., Lang, J., Prade, H.: Inconsistency management and prioritized Syntax-Based entailment. In: Proceedings of the 13th International Joint Conference on Artificial Intelligence, pp. 640–647. Morgan Kaufmann, Chambery (1993)
18. Gardenfors, P., Makinson, D.: Nonmonotonic inference based on expectations. Artificial Intelligence Journal **65**, 197–245 (1994)
19. Everaere, P., Konieczny, S., Marquis, P.: Disjunctive merging: Quota and Gmin merging operators. Artificial Intelligence Journal **174**, 824–849 (2010)
20. Konieczny, S., Pérez, R.P.: On the frontier between arbitration and majority. In: Proceedings of the Eights International Conference on Principles and Knowledge Representation and Reasoning, pp. 109–120. Toulouse, Morgan Kaufmann (2002)
21. Jinxin, L., Mendelzon, A.O.: Knowledge base merging by majority. In: Dynamic Worlds from the Frame Problem to Knowledge Management, pp. 195–218. Springer Netherlands, Toulouse (2002)
22. Revesz, P.Z.: On the semantics of arbitration. Int. J. Autom. Comput. **7**, 133–160 (1997)
23. Baral, C., Kraus, S., Minker, J.: Combining multiple knowledge bases. IEEE Trans. Knowl. Data Eng. **3**, 208–220 (1991)
24. Bloch, I., Hunter, A., Appriou, A., Ayoun, A., Benferhat, S., Besnard, P., Cholvy, L., Cooke, R.M., Cuppens, F., Dubois, D., Fargier, H., Grabisch, M., Kruse, R., Lang, J., Moral, S., Prade, H., Saffiotti, A., Smets, P., Sossai, C.: Fusion: General concepts and characteristics. Int. J. Intel. Syst. **16**, 1107–1134 (2001)
25. Benferhat, S., Bouraoui, Z., Loukil, Z.: Min-Based fusion of possibilistic DL-lite knowledge bases. In: International Conferences on Web Intelligence, pp. 23–28. IEEE Computer Society, Atlanta (2013)
26. Moguillansky, M., Falappa, M.A.: A Non-Monotonic description logics model for merging terminologies, inteligencia artificial. Revista Iberoamericana de Inteligencia Artificial **11**, 77–88 (2007)
27. Halevy, A.Y.: Answering queries using views: a survey. Very Large Data Bases Journal **10**, 270–294 (2001)

28. Pottinger, R., Halevy, A.Y.: Minicon: A scalable algorithm for answering queries using views. Very Large Data Bases Journal **10**, 182–198 (2001)
29. Pottinger, R., Levy, A.Y.: A scalable algorithm for answering queries using views. In: Proceedings of 26th International Conference on Very Large Data Bases, pp. 484–495. Morgan Kaufmann, Cairo (2000)
30. Arenas, M., Bertossi, L.E., Chomicki, J.: Consistent query answers in inconsistent databases. In: Proceedings of the Eighteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, pp. 68–79. ACM Press, Philadelphia (1999)
31. Bertossi, L.E.: Database Repairing and Consistent Query Answering. Morgan & Claypool Publishers (2011)
32. Chomicki, J.: Consistent query answering: five easy pieces. In: 11Th International Conference of Database Theory, pp. 1–17. Springer, Barcelona (2007)
33. Decker, H.: Basic causes for the inconsistency tolerance of query answering and integrity checking. In: Database and Expert Systems Applications, DEXA, International Workshops, pp. 318–322. IEEE Computer Society, Bilbao (2010)
34. ten Cate, B., Halpert, R.L., Kolaitis, P.G.: Practical query answering in data exchange under Inconsistency-Tolerant semantics. In: Proceedings of the 19th International Conference on Extending Database Technology, pp. 233–244. OpenProceedings.org, Bordeaux (2016)
35. Baral, C., Kraus, S., Minker, J., Subrahmanian, V.S.: Combining knowledge bases consisting of first order theories. In: 6Th International Symposium of Methodologies for Intelligent Systems ISMIS, pp. 92–101. Springer, Charlotte (1991)
36. Nebel, B.: Base revision operations and schemes: semantics, representation and complexity. In: ECAI, pp. 341–345 (1994)
37. Benferhat, S., Dubois, D., Prade, H.: Some syntactic approaches to the handling of inconsistent knowledge bases: a comparative study part 1: the flat case. Stud. Logica. **58**, 17–45 (1997)
38. Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., Savo, D.F.: Inconsistency-tolerant semantics for description logics. In: Fourth International Conference of Web Reasoning and Rule Systems RR, pp. 103–117. Springer, Bressanone/Brixen (2010)
39. Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., Savo, D.F.: Inconsistency-tolerant query answering in ontology-based data access. J. Web Semantics **33**, 3–29 (2015)
40. Bienvenu, M., Rosati, R.: Tractable approximations of consistent query answering for robust ontology-based data access. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence, pp. 775–781. IJCAI/AAAI, Beijing (2013)
41. Benferhat, S., Bouraoui, Z., Croitoru, M., Papini, O., Tabia, K.: Non-Objection Inference for Inconsistency-Tolerant query answering. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence IJCAI, pp. 3684–3690. IJCAI/AAAI Press, New York (2016)
42. Baget, J., Benferhat, S., Bouraoui, Z., Croitoru, M., Mugnier, M., Papini, O., Rocher, S., Tabia, K.: A general Modifier-Based framework for Inconsistency-Tolerant query answering. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference KR, pp. 513–516. AAAI Press, Cape Town (2016)
43. Grau, B.C., Motik, B.: Reasoning over Ontologies with Hidden content: The Import-by-Query Approach, coRR - Computing Research Repository. Volume arXiv:abs/1401.5853 (2014)
44. Fokoue, A., Meneguzzi, F., Sensoy, M., Pan, J.Z.: Querying linked ontological data through distributed summarization. In: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence. AAAI Press, Toronto (2012)
45. Boughammoura, R., Omri, M.N.: Querying deep web data bases without accessing to data. In: 13Th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, pp. 597–603. IEEE, Guilin (2017)
46. Boughammoura, R., Omri, M.N., Hlaoua, L.: Information retrieval from deep web based on visual query interpretation. Int. J. Inform. Res. Rev. IJIRR **2**, 45–59 (2012)
47. Boughammoura, R., Hlaoua, L., Omri, M.N.: G-Form: a Collaborative Design Approach to Regard Deep Web Form as Galaxy of Concepts. In: 12Th International Conference of Cooperative Design, Visualization, and Engineering, pp. 170–174. Springer, Mallorca (2015)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3. **Scientific Journal of King Faisal University Basic and Applied Sciences**

Article [J3] : Lexicographic repair under querying prioritised dl-lite knowledge bases.

Auteurs : Telli Abdelmoutia**, Hamdi Ghassen** and Omri Mohamed Nazih

Journal : Scientific Journal of King Faisal University Basic and Applied Sciences

Etat : Publié

Quartile : Q3

SJR : 0.14

ISSN : 16580311

KFU
جامعة الملك فيصل
KING FAISAL UNIVERSITY
Scientific Journal For King Faisal University
Search...
Scientific Journal / Issues Years Basic and Applied Sciences / Issues / Volume vol.22-Issue No.1-June 2021 (1442)
Scientific Journal of King Faisal University / Basic and Applied Sciences
Lexicographic Repair Under Querying Prioritised DL-Lite Knowledge Bases
(Telli Abdelmoutia , Hamdi Ghassen and Omri Mohamed Nazih)

---------- Forwarded message ---------
De : **Adel Abou-Salama** <onbehalfof@manuscriptcentral.com>
Date: jeu. 7 janv. 2021 à 06:45
Subject: The Scientific Journal of King Faisal University: Basic and Applied Sciences - Decision on Manuscript ID SJKF
لجامعة الملك فيصل، رقم القيد
To: <a.telli@univ-biskra.dz>

07-Jan-2021

سعادة Dear Dr. Abdelmoutia:

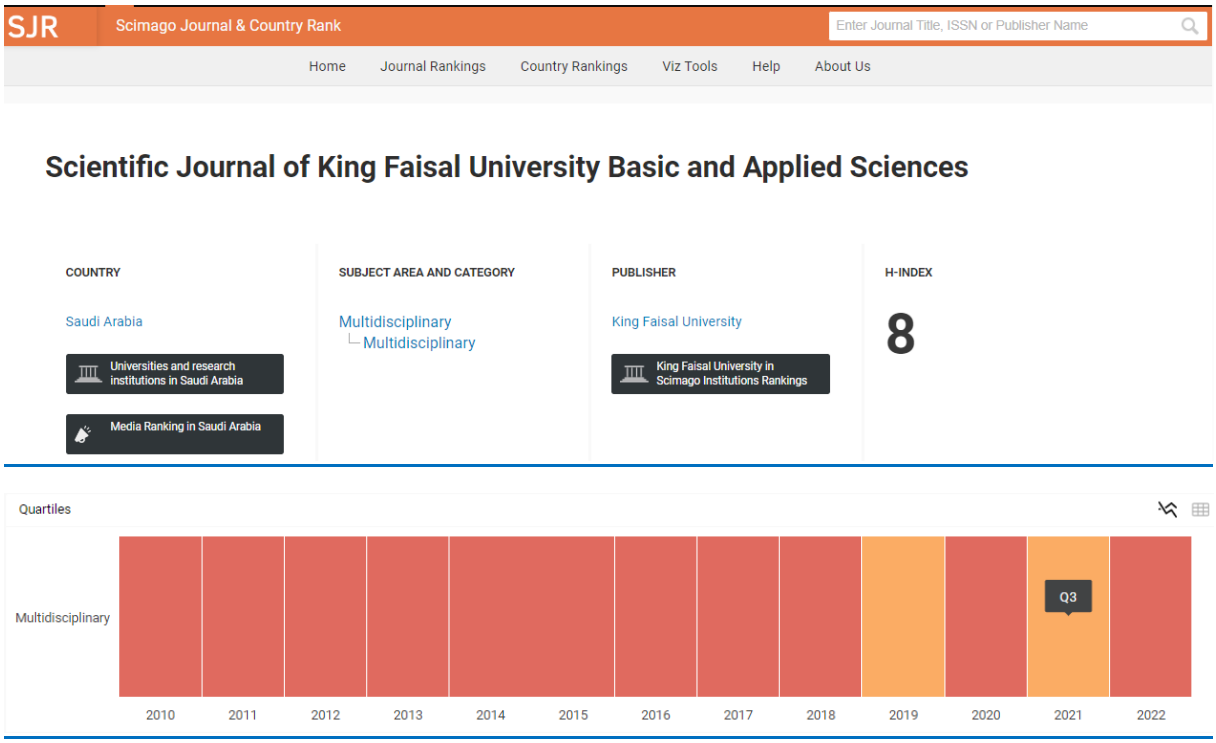
It is a pleasure to inform you that, based on the recommendations of the editorial board and reviewers, your manuscript Prioritized DL-Lite Knowledge Bases" will be accepted in the The Scientific Journal of King Faisal University: Basic and technical corrections in the manuscript. The Editor-in-Chief will contact you shortly and inform you of these corrections.

يسرنا إعلام سعادتكم بأنه بناء على توصية أصحاب السعادة المحررين وهيئة التحرير، فإنه سيتم قبول ورقتكم المعنون لها
Lexicographic Repair Under Querying Prioritized DL-Lite Knowledge Bases
في حال إكمال بعض الإجراءات الإدارية
وعليه، سيتواصل سعادة رئيس هيئة التحرير مع سعادتكم قريباً من أجل إكمال هذه الإجراءات لإصدار خطاب القبول

The comments of the reviewer(s) who reviewed your manuscript are included at the foot of this letter.

تجدون أدناه - إن وجدت - ملاحظات المحكمين

Thank you for your fine contribution. On behalf of the Editors of the The Scientific Journal of King Faisal University: Ba continued contributions to the Journal.
نشكر لكم نشركم لإنتاجكم العلمي معنا، ونتطلع لمزيد من التعاون


SJR
Scimago Journal & Country Rank
Enter Journal Title, ISSN or Publisher Name
Home
Journal Rankings
Country Rankings
Viz Tools
Help
About Us
Scientific Journal of King Faisal University Basic and Applied Sciences
COUNTRY
Saudi Arabia
Universities and research institutions in Saudi Arabia
Media Ranking in Saudi Arabia
SUBJECT AREA AND CATEGORY
Multidisciplinary
Multidisciplinary
PUBLISHER
King Faisal University
King Faisal University in Scimago Institutions Rankings
H-INDEX
8
Quartiles
Multidisciplinary
Q3
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022

The Scientific Journal of King Faisal University

تلي عبد المطيع [1] وحامدي غسان [2] وعمري محمد نزيه [2]

[1] قسم الإعلام الآلي، جامعة محمد خيضر، بسكرة، الجزائر

[2] مخبر البحث نمذجة أنظمة الاستدلال الآلي ، جامعة سوسة، سوسة، تونس

Telli Abdelmoutia [1], Hamdi Ghassen [2] and Omri Mohamed Nazih [2]

[1] Department of Computer Science, Biskra University, Biskra, Algeria

[2] MARS Research Laboratory, University of Sousse, Sousse, Tunisia

نهدي سعادتكم أعذب التحايا وأعطرها، ويسرنا إفادتكم بأنه وبناء على توصيات أصحاب السعادة أعضاء هيئة تحرير ومحكمي المجلة العلمية لجامعة الملك فيصل، فإن ورقتكم قبلت ونشرت وفقاً للبيانات أدناه. وبهذه المناسبة، نشكر لكم هذه المساهمة العلمية، ونرجو استمرار التواصل العلمي مستقبلاً، ونهيب بكم مشاركة الورقة مع المهتمين من معارفكم عبر قنوات التواصل المتعددة.

Having thoroughly read your manuscript by our editorial board members and reviewers, I am pleased to inform you that it has been accepted and published. You are encouraged to share it with colleagues and contacts via social networks. We appreciate that you have chosen our Journal and hope to receive more of your writings.

Journal in Arabic	المجلة العلمية لجامعة الملك فيصل: العلوم الأساسية والتطبيقية	المجلة بالعربي
Journal in English	The Scientific Journal of King Faisal University: Basic and Applied Sciences	المجلة بالإنجليزي
Description in Arabic	مجلة علمية محكمة نصف سنوية وذات "الوصول المفتوح"	وصف بالعربي
Description in English	An open-access refereed journal published bi-annually	وصف بالإنجليزي
Location (in Arabic)	مبنى 30، جامعة الملك فيصل، الأحساء، المملكة العربية السعودية	الموقع الحي (عربي)
Location (in English)	Building 30, King Faisal University, Al Ahsa, Saudi Arabia	الموقع الحي (انجليزي)
Website	https://www.kfu.edu.sa/ar/departments/sjournal/	الموقع الإلكتروني
ISSN	1658-0311	ردمد مطبوع
E-ISSN	1441/196	ردمد إلكتروني
Established in	2000	تاريخ إطلاق المجلة
Arabic Title	الإصلاح الخطي للاستعلام عن قواعد معرفة ذات الأولوية بالمنطق الخفيف	عنوان الورقة (عربي)
English Title	Lexicographic Repair Under Querying Prioritised DL-Lite Knowledge Bases	عنوان الورقة(انجليزي)
Issue No.	22	رقم العدد
Volume No.	1	رقم المجلد
Year	2021	السنة
DOI	10.37575/b/sci/0054	الرقم التعريفي
Received	16/11/2020	الاستقبال
Accepted	20/01/2021	القبول
Published Online	20/01/2021	النشر الإلكتروني
Article Link	https://doi.org/10.37575/b/sci/0054	رابط البحث
QR		كيو آر
Note	سيكون الرابط والكود فعال خلال 4 أيام عمل	تنويه
Note	The link and code will be available within 4 working days.	تنويه
Assigned to an Issue	بمجرد نشر الورقة إلكترونياً فإنها تحال فوراً إلى قسم "العدد في طور الاكتمال". يكتمل هذا العدد بتاريخ 01/06/2021	الإحالة لعدد
Assigned to an Issue	Once the manuscript is published online, it is immediately assigned to an 'in-progress issue'. This issue will be complete on 01/06/2021	الإحالة لعدد
Issue Link	https://services.kfu.edu.sa/ScientificJournal/ar/Home/Issues?sId=2&year=2021	رابط العدد
Citation	Abdelmoutia, T., Ghassen, H. and Nazih, O.M. (2021). Al'iislah almuejimiu dimn alaistielam ean qawaeid almaerifat DL-Lite dhat al'awlawia 'Lexicographic repair under querying prioritised DL-Lite knowledge bases'. *The Scientific Journal of King Faisal University: Basic and Applied Sciences*, **22**(1), 124–30. DOI: 10.37575/b/sci/0054	الإحالة
Citation	عبدالمطيع، تلي و غسان، حامدي و نزيه، عمري، محمد. (2021). الإصلاح الخطي للاستعلام عن قواعد معرفة ذات الأولوية بالمنطق الخفيف. *المجلة العلمية لجامعة الملك فيصل: العلوم الأساسية والتطبيقية*، **22**(1)، 124-130.	الإحالة
Author Satisfaction Survey	https://forms.gle/GLoT4si6184p81TY9	استبيان رضا المؤلف
Hashtag	#المجلة_العلمية_لجامعة_الملك_فيصل #scientific_journal_of_king_faisal_university	هاش تاق

Dr Abdulrahman Al Lily

د. عبدالرحمن الليلي

Editor-in-Chief

رئيس هيئة التحرير

scijkfu@kfu.edu.sa

المجلة العلمية لجامعة الملك فيصل
The Scientific Journal of King Faisal University

العلوم الأساسية والتطبيقية
Basic and Applied Sciences

Lexicographic Repair Under Querying Prioritised DL-Lite Knowledge Bases

Telli Abdelmoutia [1], Hamdi Ghassen [2] and Omri Mohamed Nazih [2]
[1] Department of Computer Science, Biskra University, Biskra, Algeria
[2] MARS Research Laboratory, University of Sousse, Sousse, Tunisia

الإصلاح الخطي للاستعلام عن قواعد معرفة ذات الأولوية بالمنطق الخفيف

تلي عبد المطيع [1] و حامدي غسان [2] و عمري محمد نزيه [2]
[1] قسم الإعلام الآلي، جامعة محمد خيضر، بسكرة، الجزائر
[2] مخبر البحث نمذجة أنظمة الاستدلال الآلي ، جامعة سوسة، سوسة، تونس

LINK الرابط	RECEIVED الاستقبال	ACCEPTED القبول	PUBLISHED ONLINE النشر الإلكتروني	ASSIGNED TO AN ISSUE الإحالة لعدد
https://doi.org/10.37575/b/sci/0054	16/11/2020	20/01/2021	20/01/2021	01/06/2021
NO. OF WORDS عدد الكلمات	**NO. OF PAGES عدد الصفحات**	**YEAR سنة العدد**	**VOLUME رقم المجلد**	**ISSUE رقم العدد**
6155	7	2021	22	1

ABSTRACT

This article discusses the issue of inconsistency in responses from various DL-Lite knowledge bases. This inconsistency problem is at the origin of several sources of assertions with different levels of reliability. The various solutions proposed in the literature that have to do with retrieving an exhaustive and coherent list of responses are not satisfactory from the point of view of reliability and performance. The solution that we present to solve this problem is articulated around two phases: the first phase consists of interrogating the different knowledge bases to retrieve all of the possible answers, which may be inconsistent and/or contradictory, and the second phase consists in repairing these inconsistencies and/or contradictions. To do this, we propose an approach based on three algorithms that we developed in this framework: a first algorithm for non-defeat repair, a second algorithm for lexicographic repair and a third algorithm for non-defeat repair based on lexicography of possible inconsistent responses. The experimental study carried out on the different data collections, as well as the analysis of the results obtained, confirm the performance of our approach as well as its efficiency in regards to productivity and complexity in terms of execution time.

الملخص

تناقش هذه المقالة مشكلة عدم الاتساق في الاستجابات من مختلف قواعد المعرفة DL-Lite ذات الأولوية. أصل مشكلة التناقض هي المصادر المتعددة للتأكيدات بمستويات مختلفة من الموثوقية. الحلول المختلفة المقترحة في الأدبيات من أجل استعادة قائمة شاملة ومتماسكة من الإجابات ليس مرضية من وجهة نظر الموثوقية والأداء. يعتمد الحل الذي نقدمه لحل هذه المشكلة حول مرحلتين: تتكون المرحلة الأولى، أولاً وقبل كل شيء، استجواب قواعد المعرفة المختلفة لاسترداد جميع الإجابات المحتملة التي قد تكون غير متسقة و/ أو متناقضة وتتكون المرحلة الثانية من إصلاح هذا التناقض و/ أو عدم الاتساق. للقيام بذلك، اقترحنا نهجًا يعتمد على ثلاث خوارزميات قمنا بتطويرها في هذا الإطار: خوارزمية أولى للإصلاح غير المهزوم، وخوارزمية ثانية للإصلاح الخطي وخوارزمية ثالثة للإصلاح الخطي غير المهزوم للإجابات غير المتسقة الممكنة. إن الدراسة التجريبية التي أجريت على مجموعات البيانات المختلفة المعتبرة، وكذلك تحليل النتائج التي تم الحصول عليها تؤكد أداء نهجنا وكذلك كفاءته مع الأخذ في الاعتبار الإنتاجية والتعقيد من حيث وقت التنفيذ.

KEYWORDS
الكلمات المفتاحية
Description logics, conjunctive query, answer profile, inconsistent information
المنطق الوصفي، الاستعلام المترابط، الملف الذاتي للإجابة، معلومات غير متسقة

CITATION
الإحالة
Abdelmoutia, T., Ghassen, H. and Nazih, O.M. (2021). Al'iislah almuejimiu dimn alaistielam ean qawaeid almaerifat DL-Lite dhat al'awlawia 'Lexicographic repair under querying prioritised DL-Lite knowledge bases'. *The Scientific Journal of King Faisal University: Basic and Applied Sciences*, **22**(1), 124–30. DOI: 10.37575/b/sci/0054
عبدالمطيع، تلي و غسان، حامدي و نزيه، عمري، محمد. (2021). الإصلاح الخطي للاستعلام عن قواعد معرفة ذات الأولوية بالمنطق الخفيف. *المجلة العلمية لجامعة الملك فيصل: العلوم الأساسية والتطبيقية*، **22**(1)، 124-130.

1. Introduction

Description logics (DLs) are formal ontological structures for representation and reasoning. Two distinct components are based on a DL knowledge bases (KBs): a terminological basis (called TBox), representing generic knowledge, and an assertional basis (called ABox), containing facts or assertions (Artale, 2009).

Recently, there has been a particular interest in ontology-based data access (OBDA), in which using the TBox provides improved use of the ABox while interrogating it (see Lenzerini, 2011; Poggi et al., 2008).

An ontology is typically checked and validated in such a setting, whereas the assertions may be supplied in large amounts by different and inaccurate sources that may be inconsistent with the ontology, and manually verifying and validating all the assertions is always too costly. This is why reasoning in the face of inconsistency is very important in OBDA.

In an OBDA setting, most works (essentially inspired by approaches in the database field or in sentential logic (Benferhat et al., 1993, 1997; Nebel, 1994) deal with inconsistency in KBs by proposing multiple inferences, called semantics. These semantics depend on the idea of a maximally assertional fix that firmly identifies with the thought of a fix from an information base (Lembo et al., 2010) or a maximally consistent subset using propositional logic (see Brewka, 1989; Rescher and Monor, 1970). A repair of an ABox is merely an assertional subset consistent with ontology. Assertions are also provided by multiple and possibly contradictory sources with different standards of reliability in many applications. A certain origin may have various collections of conflicting information with different levels of trust. These information sets provide an inconsistent and prioritised knowledge base (ABox).

The following statements summarise the contributions of this paper:

1. The first contribution of this paper is to present a description of the inconsistency issue in the *DL-Lite* knowledge bases,
2. The second contribution tackled in this work is a review of the literature on inconsistent and prioritised *DL-Lite* knowledge base problems,
3. Finally, a novel approach is empirically developed to repair answers under inconsistent and prioritised *DL-Lite* knowledge bases, which has not been addressed in previous scientific literature.

We organise the rest of this article as follows: we give the related works in Section 2. Then, we give the needed background about prioritised DL-Lite KBs in Section 3. In Section 4, we present the basic concepts of consistent unified answer repair. In Sections 5, we show three approaches to handling an inconsistent knowledge base. We provide our experimental analysis in Section 6, and we conclude the paper in Section 7.

Corresponding Author: Telli Abdelmoutia a.telli@univ-biskra.dz ,00213541340479 المؤلف المراسل: تلي عبد المطيع

2. Related Works

There are many scientific works that have dealt with the problem of inconsistency in DL-Lite KBs, which we can summarise in seven groups, as follows:

1. The works that deal with propositional setting, (Benferhat et al., 1995) where the authors studied the function of priorities in the management of inconsistency within the framework of sentential logic by incorporating acceptable relationships with implications capable of inferring non-trivial assumptions;
2. The works that deal with different semantics, presented essentially in the works of Trivela et al. (2019), Bienvenu et al. (2019) and Dixit (2019). Trivela et al. (2019) suggest a general structure that takes into account the semantics of Intersection Closed ABox Repair (ICAR). Bienvenu et al. (2019) contribute a practical approach for computing the query answers under three semantics, ABox Repair (AR), Intersection ABox Repair (IAR) and brave semantics in the lightweight description logic *DL-Lite*$_R$. However, Dixit (2019) suggests Consistent Answering via Satisfiability (CAvSAT) as a new solution to inconsistency response problems;
3. The works that deal with standard *DL-Lite* knowledge bases (Bertossi, 2011; Bienvenu and Rossati, 2013). The authors adapted many inconsistency algorithms via some provided standards of consistency in many applications, and they studied the data complexity of conjunctive query answering under the standard *DL-Lite* knowledge bases;
4. The works that deal with prioritised *DL-Lite* knowledge bases; particularly, the works of Lembo et al. (2010, 2015), Bienvenu and Rossati (2013), Lenzerini (2011) and Benferhat et al. (2016), which compute the consistent subsets of assertions (repairs) in order to restore the consistency of *DL-Lite* knowledge bases. In addition, Benferhat et al. (2015, 2016) propose a new non-objection inference relation based on the option of only one preferred repair and discuss the complexity of computation inference on ideal repair. Additionally, Telli et al. (2017) propose polynomial strategies for finding a unification repair consistent under all the *DL-Lite* knowledge bases;
5. The work that deals with OBDA settings, such as Bienvenu et al. (2014)'s, which is one of few works in this context and which focuses on inference with prioritised *DL-Lite* knowledge bases in OBDA settings;
6. The works that do not deal directly with KBs, such as the works of Hamdi et al. (2018), Boughammoura et al. (2012, 2015) and Boughammoura and Omri (2017), in which the authors studied several query answering strategies and proposed new approaches, querying responses from hidden datasets in which *DL-Lite* describes KB;
7. The works that deal with answers querying, such as the works of Artale et al. (2009) and Staworko et al. (2012), in which the authors explore the principle of priority answering inconsistent *DL-Lite* knowledge bases by using user preferences to limit the set of repairs down to a set of preferred repairs.

Our work belongs in a group dealing with answers querying under prioritised DL-Lite knowledge bases. For this purpose, an algorithm starts by querying each stratum of ABox in order to provide us with all the possible response sets. After that, once the response sets are consistent, no repair is applied. Otherwise, the algorithm repairs the response sets.

In addition, this paper uses DL-Lite Du and Shen (2013) due to its efficiency in conjunctive query answering and computing contradictory knowledge. Also, we relied only on conjunctive query, as it includes all basic queries, and many first-order queries can be written as conjunctive queries.

3. *DL-Lite* Knowledge Base

The DL-Lite family (Artale et al., 2009; Poggi et al., 2008) is included in OWL2 QL syntax. The knowledge representation format for DL-Lite is as follows: NC is a set of atomic concepts, NR is a set of atomic roles and NI is a set of individuals or assertions. We consider three connectors '$\neg$', '$\exists$' and '$^-$', which are used to describe complex concepts and complex roles as follows (Artale et al., 2009):

$R_1 \rightarrow R_2$ or $R_1 \rightarrow R_1^-$

$E \rightarrow R_1$ or $E \rightarrow \neg R_1$

$Bc \rightarrow Dc$ or $B \rightarrow \exists R_1$

$Cc \rightarrow Bc$ or $Cc \rightarrow \neg Bc$

such that Dc represents an atomic concept, R1 represents an atomic role and R1- represents the inverse of R1. However, Bc represents basic concept, Cc represents complex concept, R1 represents basic role and E represents complex role. A DL-Litecore knowledge base K is a pair K = ⟨T, A⟩ (Benferhat et al., 1997). T = TBox is made up of a finite set of inclusion axioms between concepts of the form B ⊆ C or B ⊆ ¬C. A = ABox contains the finite set of assertions (facts) of atomic concepts and roles of the form D(a) and P(a, b).

The DL-LiteF language extends DL-Litecore with the capability of functional specification on roles or their inverses of the form (functR). The DL-LiteR language extends DL-Litecore with the ability to specify inclusion axioms between roles in TBox of the form R ⊆ E. Note that DL-Lite language does not use connective or disjunctive operators. However, a logical transformation makes it possible to obtain conjunctions and disjunctions as follows:

- A conjunction of the form $B \subseteq C \cap D$ is equivalent to the pair of inclusion axioms $B \subseteq C$ and $B \subseteq D$;
- A disjunction of the form $C \cup D \subseteq B$ is equivalent to the pair of inclusion axioms $C \subseteq B$ and $D \subseteq B$.

Note that all DL-Lite knowledge bases can be written as a First-Order Logic (FOL) knowledge base.

In addition, we share the semantics of DL-Lite knowledge bases. A semantics is an interpretation I=(Δ^I,$\cdot^I$) from a non-empty domain Δ^I to interpret function$\cdot^I$ such that:

{∀ x ∈ NI , ∃ xI ∈ Δ^I, ∀ C ∈ NC , ∃ CI ⊆ Δ^I and ∀ R ∈ NR and ∃ R ⊆ Δ^IxΔ^I}.

Note that a DL-Lite knowledge base is inconsistent if it does not admit any model.

3.1. Prioritised Profile DL-Lite Knowledge Base:

We claim that the prioritised profile DL-Lite knowledge base $K_P = < T, P_S >$ for all sets of prioritised ABoxes such that T is a flat (standard) DL-Lite TBox, and Ps = {L1,...,Lm} is a prioritised ABox profile, where Li is a layer (stratum) i, which includes a list of assertions that have the same level of priority, and $\forall\, j > i$ Li is more important than Lj.

3.2. Conjunctive Query:

We claim that the query Q={(v)| f(v) is a First-Order Logic formula when:

- *(v)=(v$_1$,...,v$_n$)*: free variables,
- *n*: the arity of *Q* and atoms of *f(v)*,
- *f(v): D(t$_i$)* or *R(t$_i$; t$_j$)*, ∀ *D* ∈ N_C and ∀ *R* ∈ N_R and *t$_i$, t$_j$* are terms.

Note that f(v), of the form ∃ (w).conj(v; w), and w are existentially quantified variables, and conj(x; y) is a conjunction of atoms of the form D(ti) or P (ti; tj). Q is said to be a conjunctive query (CQ). An answer to a CQ Q(x) ← conj(x; y) over $K =< T, A >$ is a non-empty set of tuples s = (s1,..., sk) ∈ NI x...x NI such that $< T, A > \models Q(s)$.

Now, let Q(x) be a conjunctive query. We consider SPs = {S1 ,..., Sm} a set of responses about Q(x) under Ps, and Si = s ∈ NI x...x NI : $< T, A > \models Q(s)$ with certainty, when there is no answer to the query Q(x) with respect to Li, Si = ∅.

<u>Example 1</u>: We consider $K =< T, P_S >$ to be a prioritised DL-Lite

knowledge base.

Thus, we have

T ={C1 ⊆ ¬C2, C2 ⊆ ¬ C3, R1 ⊆ R2 and Ps ={L1, L2, L3},

where

L1 = {C2(a), R1(a, c), C1(a)},

L2 = {C1(b), R1(b, c), C3(e), R2(e, c)} and

L3 = {C3(b), R1(b, c), C3(a), R1(a, c)}.

We consider also the following conjunctive query Q, which requires any individual ν have a relationship to c through the role R1: Q(ν) = ∃ ν: R1(ν,c). The list of responses QPs to this query are

QL1 = {C1(a), C2(a)},

QL2 = {C1(b)} and

QL3 = {C3(b), C3(a)}.

4. Consistent Unified Answers Reasoning

In this section, we give a brief refresher, proposed by Benferhat et al. (1992). Then, we will use these concepts to present our approach.

The Conflict Answer Sets (CA⊆(QPs)) represent a minimal inconsistent subset CA of the assertions associated with SPs such that CA is inconsistent. Hence, ∀ α ∈ CA, CA \{α} is consistent with respect to TBox.

The Free Answer Sets (free(QPs)) represent the subset of facts ∈ QPs that are not included with the Conflict Answer Sets CA in QPs with respect to TBox . We say that an assertion β ∈ QPs is free if ∀α ∈ C(QPs): $\beta \notin \alpha$. The free individual in a propositional logic context has been previously suggested by Lutz (2013).

The answers repair RA⊆ (QL1,...,QLm) is a consistent subset, and it is denoted by MARA (maximally inclusion-based answers repair of QPs) if:

<T, RA> is consistent and

∀RA ⊆ (QL1, ..., QLm): RA ∉ R'A, R'A is inconsistent.

This definition of MARA is similar to that defined in Lembo et al. (2010).

We denote MARA(QPs) by the set of MARA of QPs with respect to T. Inconsistency in flat DL-Lite KBs can be accomplished using the principle of answers repair by applying standard request answers using the entire set of answers repair (AR-entailment (Lembo et al., 2010) or using only one answers repair.

The preferred inclusion-based answers repair RARA(QPs) is the extension of the MARA definition, when the DL-Lite ABox is prioritised (Bienvenu et al., 2014).

PARA(QPs) = PA1 ∪ ... ∪ PAm of QPs

such that :

∄MARA(QPs): P'A1 ∪ ... ∪ P'Am of QPs, and

if i is an integer: PAi ∈ P'Ai , ∀j=1, ..., (i-1), PAj = P'Aj.

Example 2: We continue from the previous example.

According to the definition of the Conflict Answer Sets, we have

CA(QPs) = {(C1(a); C2(a)), (C2(a); C3(a))}.

According to the definition of the Free Set Answers, we have

free(QPs) = {(C1(b) ; C3(b))}.

While, according to the definition of the preferred inclusion-based answers repair, we have

PARA(QPs) = {C1(a), C1(b), C3(b)}.

The next section discusses a repair that is used as a selection of facts to deal with inconsistent answers and propose new approaches based on a lexicographic approach. The result of the given approaches is a consistent set of corresponding answers.

5. Approaches for Repairing Answers Profile

For the rest of the paper, we consider the following settings:

- $K_P = < T, P_S >$ as a prioritised *DL-Lite* knowledge base;
- $P_s = \{L_1,...,L_m\}$ as a prioritised ABox profile;
- Q as a conjunctive query;
- $S_{Ps} = \{S_1,..., S_m\}$ as a set of answers to a query Q with respect to P_s;
- $Q_{Ps} = (Q_{L1},...,Q_{Lm})$ as a set of assertions associated with S_{Ps};
- $Q_{Li} = \{Q(s): s \in Si\}$, where Q_{Li} is a set of answers to the query Q for each L_i.

For a given repair of answers, we start by interrogating each ABox with a conjunctive query. Then, we recover answers for each level of ABox. Finally, one of the proposed approaches will be run to repair the total set of responses. Algorithm 1 presents the steps necessary to obtain consistent unified answers under a prioritised profile DL-Lite knowledge base.

Algorithm 1: Response processing from inconsistent *DL-Lite* knowledge bases

Data: A prioritised *DL-Lite* KBs $<T, (L_1,...L_m)>$
Conjunctive query Q
Result: Consistent unified answers
Interrogation of each ABox L_i by Q
Recovery answers for each ABox $<T, (Q_{L1},...Q_{Li})>$
If $<T, (Q_{L1},...Q_{Lm})>$ is consistent, **then**
Consistent unified answers are **returned**
Or
Repairing using one of the proposed approaches
Returns consistent unified answers
End.

5.1. Non-Defeated Repair of Inconsistent Answers

This new repair of inconsistent answers consists of evaluating the set of assertions associated with the answers to a given query in reference to the ABox profile proposed by Benferhat et al. (1992). The non-defeated repair of inconsistent responses is ndA(QPs) = A'1 ∪ ... ∪ A'm, where

∀i = 1... m; A'i = free(QL1 ∪ ... ∪ QLi);

namely, ndA(QPs) = free(QL1) ∪ free(QL1 ∪ QL2)∪...∪ free(QL1 ∪... ∪QLm).

The non-defeated answers repair is computed in polynomial time in DL-Lite, as it appears in Algorithm 2. It starts by initialising the set of ndA repair. Then, it computes the set of free assertions in (QL1 ∪ ... ∪ QLi), which is done in polynomial time because the free set assertions can be computed in linear time with respect to the conflict set assertions. Hence, the computation of ndA(QPs) is also done in polynomial time.

Algorithm 2: Non-Defeated Repair of Inconsistent Answers

Data: A prioritised DL-Lite KBs <T, (L1,...Lm)>
Conjunctive query Q
Result: Consistent unified answers ndA(QPs)
Interrogation of each ABox Li by Q
Recovery answers for each ABox <T, (QL1,...QLm)>

$$nd_A(Q_{Ps}) \leftarrow \phi$$

for i=1 to m do

$$nd_A(Q_{Ps}) \leftarrow nd_A(Q_{Ps}) \cup free(Q_{L1} \cup ... \cup Q_{Li})$$

return ndA(QPs)
end.

Example 3: We continue from the previous example.

We have free(QL1) = {∅}, free(QL2) = {C1(b)} and free(QL3) = {C3(b), C3(a)}.

Hence, ndA(QPs) = {C1(b); C3(b); C3(a)}.

Clearly, A query Q is said to be a ndA(QPs)-the consequence of KPs, denoted by KPs ⊨ Q, if and only if KP = <T, ndA(QPs)> ⊨Q.

5.2. Lexicographic Repair of Inconsistent Answers

In the propositional context, lexicographic inference has been commonly used by Benferhat et al. (1993). One approach to

lexicographic repair is PARLex(QPs) of inconsistent answers, which is based on the cardinality criterion instead of the set inclusion criterion.

The lexicographic repair of inconsistent answers PARLex(QPs) is defined as follows:

∀ PQps = PA1 ∪ … ∪ PAm ∈ PARA(QPs): ∄ i such that |PAi| > |LAi| and ∀j < I, |PAj| = |LAi|, where |X| represent the cardinality of the set X.

Clearly, using a lexicographic-based approach comes down to selecting among the set of repairs in PARA(QPs), particularly the ones having the maximal number of elements (See Algorithm 3).

Algorithm 3: Lexicographic Repair of Inconsistent Answers

Data: A prioritised DL-Lite KBs <T, (L1,…Lm)>
Conjunctive query Q
Result: Consistent unified answers PARLex(QPs)
Interrogation of each ABox Li by Q
Recovery answers for each ABox <T, (QL1,…QLm)>

$$PAR_{Lex}(Q_{Ps}) \leftarrow \phi$$

for i=1 to m do

$$PAR_{Lex}(Q_{Ps}) \leftarrow PAR_{Lex}(Q_{Ps}) \cup PAR(Q_{L1} \cup \ldots \cup Q_{Li})$$

return PARLex(QPs)
end.

In particular, the large number of PARLex(QPs) that can be determined from an inconsistent DL-Lite knowledge base is one of the main PAR-entailment issues. However, Algorithm 3 proceeds from the first answers set to the less preferred ones. Thus, the assertions selected do not conflict with the ones in the first answers set in order to ensure the maximally prioritised returned set with respect to lexicographic ordering. Hence, this algorithm is based on checking for inconsistency, and its computational complexity is polynomial.

Example 4: We continue from the previous example.

We have PARLex(QPs) = {C2(a), C1(b), C3(b)}. Clearly, a query Q is said to be a PARLex(QPs)-the consequence of KPs, denoted by KPs ⊨ Q, if and only if KP = <T, PARLex(QPs)> ⊨ Q.

The next subsection proposes an approach that consists of introducing a cardinality criterion instead of a set inclusion criterion and is based on non-defeated repair of inconsistent answers. We discuss lexicographic-based non-defeated repair of inconsistent answers.

5.3. Lexicographic-Based Non-Defeated Repair of Inconsistent Answers

The lexicographic-based non-defeated repair of inconsistent answers denoted by

ndLex(QPs) = L'1 ∪ … ∪ L'm is defined as follows:

∀i = 1 to m: Li = ∩ RA ∈ MARALex(L1 ∪ … ∪ Li) RA,

where MARALex(Li) = RA: RA ∈ MARA(Li), and ∄ R'A ∈ MARA(L) such that |R'A| > |RA|.

Algorithm 4: Lexicographic-Based Non-Defeated Repair of Inconsistent Answers

Data: A prioritised DL-Lite KBs <T, (L1,…Lm)>
Conjunctive query Q
Result: Consistent unified answers PARLex(QPs
Interrogation of each ABox Li by Q
Recovery answers for each ABox <T, (QL1,…QLm)>

$$nd_{Lex}(Q_{Ps}) \leftarrow \phi$$

for i=1 to m do

$$nd_{Lex}(Q_{Ps}) \leftarrow nd_{Lex}(Q_{Ps}) \cup PAR_{Lex}(Q_{Li})$$

return PARLex(QPs)
end.

According to the Algorithm 4, the main advantage of the lexicographic-based non-defeated repair of inconsistent answers approach is the production of more conclusions than the standard non-defeated repair of inconsistent answers approach, ndA(QPs). Note that this algorithm is based on the previous polynomial algorithms. Hence, it is also done in polynomial time.

Example 5: We continue from the previous example.

We have ndLex(QPs)={C1(a); C1(b); C3(b); C3(a)}.

Clearly, a query Q is said to be a ndLex(QPs)-the consequence of KPs denoted by KPs ⊨ Q, if and only if KP = <T, ndLex(QPs)> ⊨Q.

6. Experimental Analysis

This section presents an experimental analysis of the running time and productivity of our proposed approaches.

6.1. Software and Hardware Environments

We have implemented our algorithms to compute a consistent unified answer in Java programming language, Web Ontology Language Second Edition-Query Language (OWL2-QL) function syntax and a Structured Query Language Lite (SQLite) database engine for relational database manipulation. Then, we used:

- The benchmark existing on https://code.google.com/p/combo-obda;
- The *TBox* of ontology Lehigh University Benchmark ∃ (LUBM∃) *20* (Calvanese et al., 2005) and
- The Extended University Data Generator (EUDG) for generating the ABoxes.

All experiments were performed on an ASUS Sonic Master Introduction laptop Model X556QUK with an Intel (R) Core(TM) (i5) 7200 CPU @ 2.50 GHz 2.71 GHz. 4 GB DDR3 RAM. This hardware configuration is installed in a 64-bit operating system, x64 processor (Windows 10 Home).

6.2. Experimental Parameters

The theoretical foundations of this work can be found in Calvanese et al. (2005)'s document. Their work explains how exactly we modified the original LUBM data generator and ontology. It consists an evaluation of the ABox, which stores relational database (DB) queries expressed from the negative closure of the TBox to exhibit whether the KBs contains conflicting elements. This negative closure of a KBs is made of the list of all negative axioms of the form (B ⊆ ¬C), which can be derived from TBox by applying positive rules onto negative ones. The LUBM∃ 20 ontology contains the axioms presented in Table I. Our proposed approaches are based on a DL-Lite ontology parser and an SQLite database engine. They are also specifically focused on some of the following operations: checking consistency and checking conflict.

We emphasised that the process of computing the conflicts set would be realised once and for all and kept it in mind during all experimentations. After all the settings were available, we proceeded to repair the DL-Lite knowledge base using the different approaches proposed in this work.

Table 1. Ontology LUBM Information

Axioms	Size	Examples
Classes	129	FullProfessor, Faculty
Object Property	28	HasFaculty, isPartOfUniversity
Data Property	7	DataPropertyDomain(age Person)
Subclass Of	153	SubclassOf(Professor Faculty)
Disjoint Class Of	643	DisjointClassOf (Techer, Student)
Subobject Property Of	5	SubobjectPropertyOf(headOf worksFor)
Inverse Object Properties	3	InverseObjectProperties(HasSupervisor)
Disjoint Object Properties	227	DisjointObjectPropertyOf(headOf advisor)
Object Property Domain	25	ObjectPropertyDomain(advisor Person)
Object Property Range	22	ObjectPropertyRange(advisor Professor)
Data Property Domain	4	DataPropertyDomain(age Person)

6.3. Tests and Results

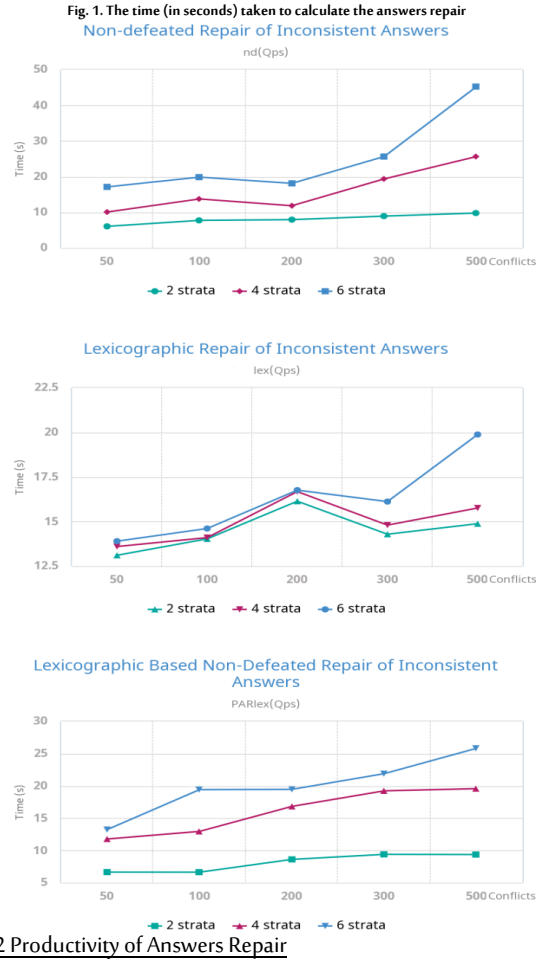
We used an SQLite engine to calculate the conflict elements and to check for inconsistencies. This allowed for efficient management of inconsistency. Then, using EUDG, we generated an ABox, divided it into 2, 4 and 6 strata, each with 50, 100, 200, 300 and 500 sets of conflict elements. For each case, we launched conjunctive query, and we collected the corresponding answers. We focused on two important features: calculation time and productivity of answers repair to evaluate our proposed algorithms. We analysed our

approaches by evaluating the results observed in the following graphs:

6.3.1. Calculation Time of Answers Repair

By calculation time of answers repair, we mean the time it took to compute our proposed algorithms of repairs. The set of graphs in Fig. 1 presents the results obtained in these experimentations.

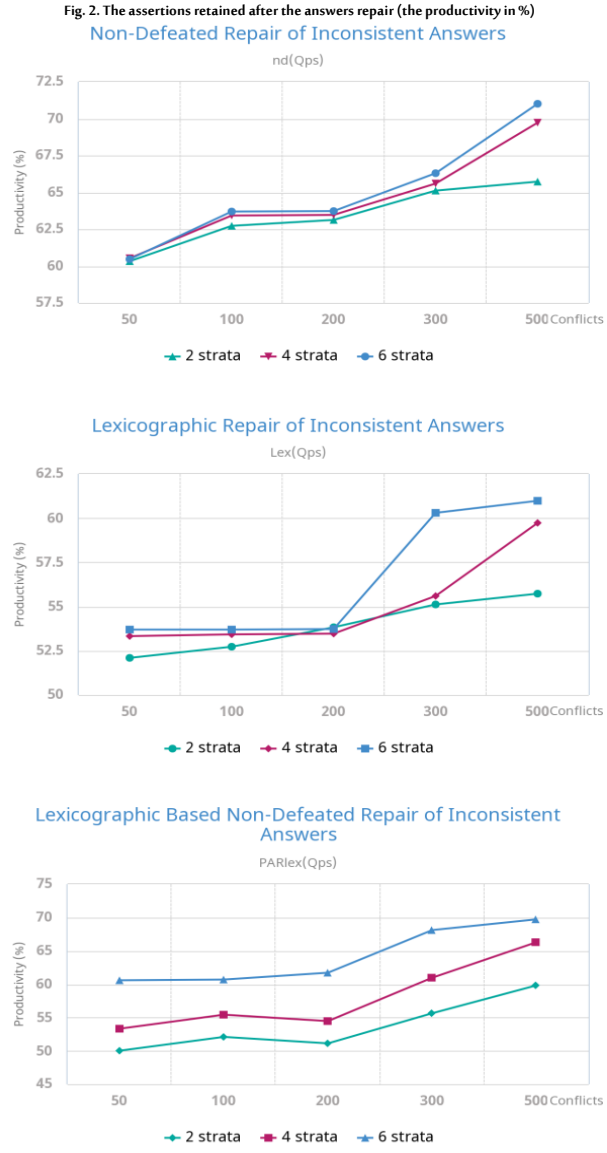
Fig. 1. The time (in seconds) taken to calculate the answers repair



6.3.2 Productivity of Answers Repair

By productivity, we mean all assertions that were retained from the answers that restore the answers' consistency. The results obtained in this experiment are presented in Fig. 2.

Fig. 2. The assertions retained after the answers repair (the productivity in %)



6.4. Analysis and Evaluation of Results

According to Fig 1, the running time to calculate the lexicographic-based answers repair using the non-defeated approach was greater than the lexicographic and non-defeated approaches. Thus, the running time was affected by increasing the number of conflicts and the number of strata.

Similarly, the productivity of the answers repair shown in Fig 2 proves that the lexicographic-based non-defeated answers repair algorithm was more productive than the lexicographic and non-defeated approaches in all cases. Consequently, the productivity was affected proportionately by the size of the conflicts set and the number of strata in the ABox.

More precisely, there was a convergence between all proposed approaches to running time and productivity when the size of the conflicts set was less than 200 elements. Nevertheless, we found significant increases in running time and productivity after that.

These results proved that the productivity (number of answers returned by applying query) of an applied lexicographic base to non-defeated answers repair on prioritised KBs was more than the repairs standard approach to non-defeated answers repair.

Generally, there was a high degree of efficiency when the

correspondence between the query and the number of requested assertions was significant. We also noticed that the lexicographic-based non-defeated approach was the best performing repair approach, compared to the other two approaches.

Finally, we noted that the calculations of conflict sets had an important impact on our approaches due to having to continuously update the degree of inconsistency, which can be increased when new assertions are added. Hence, it suffices to account for the new conflicts that arise as a result of adding new assertions.

Our investigation yielded some positive results, including the development of new polynomial algorithms, such as lexicographic-based non-defeated answers repair with conjunctive query. The analysis we conducted for consistent query answering in lightweight ontologies proved that our repairs were computed incrementally, starting from the first layer until the last one.

In summary, the size of conflicting elements and the number of layers in the ABoxes are considered principal properties that directly influence the running time and the productivity of our approaches. Particularly, the lexicographic-based ndA-inference offers an important number of consistent answers compared to the ndA-inference and the Lex-inference.

7. Conclusion

In this article, we focused on the problem of inconsistent answers from prioritised DL-Lite KB. After giving the basic concepts of DL-Lite hierarchical knowledge bases and how to fix inconsistent responses of conjunctive queries, we presented an approach which relies on three algorithms in order to repair all possible answers, instead of fixing all the knowledge bases.

Our approach consisted of, first, a non-defeated repair, then a lexicographic repair and, finally, a non-defeated repair based on the lexicographic of possible incoherent responses. These different repairs are done in a polynomial time, which allowed us to reduce the operating complexity and increase productivity while remaining efficient, as shown in the paragraph which presents the experimental study.

In the future, we will expand interest through the following axes. The first axis is to compare in-depth our approach, and the main approaches in the literature, with other data collections in order to give researchers more studies and analyses about handling the problem of repairing inconsistent and/or contradictory responses in the context of priority queries in DL-Lite knowledge bases. Through this study, we hope to definitively confirm the performance and robustness of our proposed approach. As a second axis, we plan to study how to deal with missing information (conflict answer sets) in knowledge bases. And as a third axis, we will add how to repair knowledge bases starting with repairing answers.

Biographies

Telli Abdelmoutia

Department of Computer Science, Biskra University, Biskra, Algeria, 00213541340479, a.telli@univ-biskra.dz

Dr Abdelmoutia is an Algerian senior lecturer. He received his PhD degree from Biskra University in Algeria in 2018 with a Co-tutelle from Artois University in Lille (France). His current research interests concern the representation of knowledge, the management of uncertainties and the management of conflicts in priority knowledge bases. In addition, he is a member of the CRIL-CNRS Lab (Artois University – France) and the PRFU project (M'sila University – Algeria).

orcid.org/0000-0002-2907-9782.

Hamdi Ghassen

MARS Research Laboratory, University of Sousse, Sousse, Tunisia, 0021641655264, hamdighassan@gmail.com

Dr Ghassen is a Tunisian teacher. He received his PhD degree in computer science from the Higher Institute of Computer Science and Communication Technologies of Hamman Sousse in Sousse, Tunisia in December 2020. His research activity is particularly focused on knowledge representation and reasoning. Moreover, he is a member of the Modelling of Automated Reasoning Systems (MARS) Research Laboratory.

Omri Mohamed Nazih

MARS Research Laboratory, University of Sousse, Sousse, Tunisia, 0021673369502, mohamednazih.omri@eniso.u-sousse.tn

Prof. Nazih has been a Tunisian professor of computer science since 2011. He received his PhD degree from Jussieu University in Paris, France in 1994. He is a founding member of the Modelling of Automated Reasoning Systems (MARS) Research Laboratory. His group conducts research on information retrieval, Data Base and knowledge bases. He is a reviewer of many international journals, such as Information Fusion and Expert System and Application, and many international conferences: AMIA, ICNC-FSKD and AMAI.

References

Arenas, M., Bertossi, L. and Chomicki, J. (1999). Consistent query answers in inconsistent databases. *In*: *Proceedings of the Eighteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems,* Philadelphia, Pennsylvania, 31/05–02/06/1999. (pp. 68-79)

Artale, A., Calvanese, D., Kontchakov, R. and Zakharyaschev, M. (2009). The DL-Lite family and relations. *Journal of Artificial Intelligence Research*, **36**(1), 1–69.

Baget, J.F., Benferhat, S., Bouraoui, Z., Croitoru, M., Mugnier, M.L., Papini, O., Rocher, S. and Tabia, K. (2016). A general modifier-based framework for inconsistency-tolerant query answering. *In*: *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning (KR'16),* Cape Town, South Africa, 04–25/04/2016.

Baral, C., Kraus, S. and Minker, J. (1991). Combining multiple knowledge bases. *IEEE Transactions on Knowledge and Data Engineering*, **3**(2), 208–20.

Baral, C., Kraus, S., Minker, J. and Subrahmanian, V.S. (1972). Combining knowledge bases consisting of first-order analysis. *Comput. Intell.* **8**(1), 45–71.

Benferhat, S., Bouraoui, Z. and Tabia, K. (2015). How to select one preferred assertional-based repair from Inconsistent and prioritised DL-lite knowledge bases? *In*: *24th International Joint Conference on Artificial Intelligence*, Buenos Aires, Argentina, 25/07–01/08/2015.

Benferhat, S., Bouraoui, Z., Croitoru, M., Papini, O. and Tabia, K. (2016). Non-objection inference for inconsistency-tolerant query answering. *In*: *International Joint Conference on Artificial Intelligence*, New York, United States, 09–15/07/2016.

Benferhat, S., Cayrol, C., Dubois, D., Lang, J. and Prade, H. (1993). Inconsistency management and prioritised syntax-based entailment. *In*: *International Joint Conference on Artificial Intelligence*, Chambery, France, 28/08–03/09/1993.

Benferhat, S., Dubois, D. and Prade, H. (1992). Representing default rules in possibilistic logic. *In*: Proc. of *the 3rd Inter. Conf. on Principles of Knowledge Representation and Reasoning* (KR'92), Cambridge, MA, USA, 09–25/10/1992.

Benferhat, S., Dubois, D. and Prade, H. (1995). How to infer from inconsistent beliefs without revising? *In*: *International Joint Conference on Artificial Intelligence*, Montreal, Quebec, Canada, 05–20/08/1995.

Benferhat, S., Dubois, D. and Prade, H. (1997). Some syntactic approaches to the handling of inconsistent knowledge bases: A comparative study part 1: The flat case. *Studia Logica*, **58**(1), 17–45.

Bertossi, L. (2011). Database repairing and consistent query answering. *Synthesis Lectures on Data Management*, **3**(5), 1–121.

Bienvenu, M. (2012). On the complexity of consistent query answering in the presence of simple ontologies. *In*: *Twenty-Sixth AAAI Conference on Artificial Intelligence*, Toronto, Ontario, Canada, 06–22/07/2012.

Bienvenu, M. and Rosati, R. (2013). Tractable approximations of consistent query answering for robust ontology-based data access. *In*: *Twenty-Third International Joint Conference on Artificial Intelligence*, Beijing, China, 03–09/08/2013.

Bienvenu, M., Bourgaux, C. and Goasdoué, F. (2014). Querying inconsistent description logic knowledge bases under preferred repair semantics. *In*: *AAAI Conference on Artificial Intelligence*, Québec, Canada, 27–31/07/2014.

Bienvenu, M., Bourgaux, C. and Goasdoué, F. (2016). Query-driven repairing of inconsistent DL-Lite knowledge bases. *In*: *International Joint Conference on Artificial Intelligence*, New York, United States, 09–15/07/2016.

Bienvenu, M., Bourgaux, C. and Goasdoué, F. (2019). Computing and explaining query answers over inconsistent DL-Lite knowledge bases. *Journal of Artificial Intelligence Research*, **64**(n/a), 563–644.

Boughammoura, R. and Omri, M.N. (2017). Querying deep web data bases without accessing to data. *In*: *13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, Guilin, China, 29–30/09/2017.

Boughammoura, R., Hlaoua, L. and Omri, M.N. (2015). G-form: A collaborative design approach to regard deep web form as galaxy of concepts. In, *International Conference on Cooperative Design, Visualization and Engineering*, Mallorca, Spain, 3–20/09/2015.

Boughammoura, R., Omri, M.N. and Hlaoua, L. (2012). Information retrieval from deep web based on visual query interpretation. *International Journal of Information Retrieval Research* (IJIRR), **2**(4), 45–59.

Brewka, G. (1989). Preferred subtheories: An extended logical framework for default reasoning. *In*: *International Joint Conference on Artificial Intelligence*, Detroit, MI, USA, 05–20/08/1989.

Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M. and Rosati, R. (2005). DL-Lite: Tractable description logics for ontologies. *In*: *AAAI Conference on Artificial Intelligence*, Pittsburgh, Pennsylvania, 09–13/07/2005.

Chomicki, J. (2007). Consistent query answering: Five easy pieces. In: *International Conference on Database Theory*, Barcelona, Spain, 02–10/01/2007.

Dixit, A.A. (2019). CAvSAT: A system for query answering over inconsistent databases. *In*: *Proceedings of the 2019 International Conference on Management of Data*, Amsterdam, The Netherlands, 30/05–05/06/2019.

Du, J., Qi, G. and Shen, Y.D. (2013). Weight-based consistent query answering over inconsistent SHIQ knowledge bases. *Knowledge and Information Systems*, **34**(2), 335–71.

Hamdi, G., Omri, M.N., Benferhat, S., Bouraoui, Z. and Papini, O. (2020). Query answering dl-lite knowledge bases from hidden datasets. *Annals of Mathematics and Artificial Intelligence*. DOI: 10.1007/s10472-020-09714-2

Hamdi, G., Telli, A. and Omri, M.N., (2020). Querying of several DL-Lite knowledge bases from various information sources-based polynomial response unification approach. *Journal of King Saud University-Computer and Information Sciences*. DOI: 10.1016/j.jksuci.2020.06.002

Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M. and Savo, D. F. (2010). Inconsistency-tolerant semantics for description logics. *In*: *International Conference on Web Reasoning and Rule Systems*, Bressanone, Italy, 23–24/09/2010.

Lenzerini, M. (2011, October 24–28). Ontology-based data management. *In*: *Proceedings of the 20th ACM International Conference On Information And Knowledge Management*, Glasgow, Scotland, UK.

Lukasiewicz, T., Martinez, M.V. and Simari, G.I. (2012). Inconsistency handling in datalog+/-ontologies. *In*: *Proceedings of the 20th European Conference on Artificial Intelligence*, Montpellier, France, 27–31/08/2012.

Lukasiewicz, T., Martinez, M. V., Pieris, A. and Simari, G. I. (2015). From classical to consistent query answering under existential rules. *In*: *Proceedings of the AAAI Conference on Artificial Intelligenc*e, Austin Texas, USA, 25–30/01/2015.

Lutz, C., Seylan, I., Toman, D. and Wolter, F. (2013). The combined approach to OBDA: Taming role hierarchies using filters. *In*: *International Semantic Web Conference*, Sydney, Australia, 21–25/10/2013.

Martinez, M.V., Parisi, F., Pugliese, A., Simari, G.I. and Subrahmanian, V.S. (2008). Inconsistency management policies. *In*: *Principles of Knowledge Representation and Reasoning*, Sydney, Australia, 16–19/09/2008.

Nebel, B. (1994). Base revision operations and schemes: Semantics, representation, and complexity. *In*: *11th European Conference on Artificial Intelligence*, Amsterdam, Netherlands, 8–12/08/1994.

Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M. and Rosati, R. (2008). Linking data to ontologies. *Journal on Data Semantics X*, **10**(n/a), 133–73.

Rescher, N. and Manor, R. (1970). On inference from inconsistent premisses. *Theory and Decision*, **1**(2), 179–217.

Staworko, S., Chomicki, J. and Marcinkowski, J. (2012). Prioritised repairing and consistent query answering in relational databases. *Annals of Mathematics and Artificial Intelligence*, **64**(2-3), 209–46.

Telli, A., Benferhat, S., Bourahla, M., Bouraoui, Z. and Tabia, K. (2017). Polynomial algorithms for computing a single preferred assertional-based repair. *KI-Künstliche Intelligenz*, **31**(1), 15–30.

Trivela, D., Stoilos, G. and Vassalos, V. (2019). Query rewriting for DL ontologies under the ICAR semantics. *In*: *International Joint Conference on Rules and Reasoning*, Macao, China, 10–16/08/2019.

II. Articles soumis dans des Journaux Internationaux

1. Social Network Analysis and Mining

Article [J1] : Personalization of OLAP queries for hierarchical visualization under constraints.

Auteurs : **Hamdi Ghassen and** Omri Mohamed Nazih

Journal : Social Network Analysis and Mining

Etat : Under Review

Quartile : Q2

SJR : 0.61

ISSN: 1869-5469

Overview

Social Network Analysis and Mining is a multidisciplinary journal focusing on theoretical and experimental work related to social network analysis and mining.

- Serves a wide range of researchers from computer science, network science, social sciences, mathematical sciences, medical, biological, financial, management, and political sciences.
- 92% of authors who answered a survey reported they would definitely or probably publish in the journal again.
- It covers a broad spectrum of topics from social sciences, mathematics, statistics, physics, network science, and computer science.
- Aims to create a platform that encourages the integration of varied techniques in the analysis and mining of social networks.

This is a transformative journal, you may have access to funding.

Editor-in-Chief
Reda Alhajj

Impact factor
2.8 (2022)

5 year impact factor
2.7 (2022)

Submission to first decision (median)
10 days

Downloads
323,357 (2022)

You are the author of this submission

Personalization of OLAP queries for hierarchical visualization under constraints

PEER REVIEW TIMELINE

Follow the progress of your submission.

This submission is **UNDER REVIEW** at *Social Network Analysis and Mining*

Your manuscript is under review at *Social Network Analysis and Mining*. Get the most recent updates on its progress below.

Version 1 (private) received 23 Mar, 2023

- Reviewers agreed *On 04 Apr, 2023*
- Reviewers invited *Invitations sent on 27 Mar, 2023*
- Editor assigned *On 26 Mar, 2023*
- Submission checks complete *On 24 Mar, 2023*
- First submitted to *Social Network Analysis and Mining* *On 23 Mar, 2023*

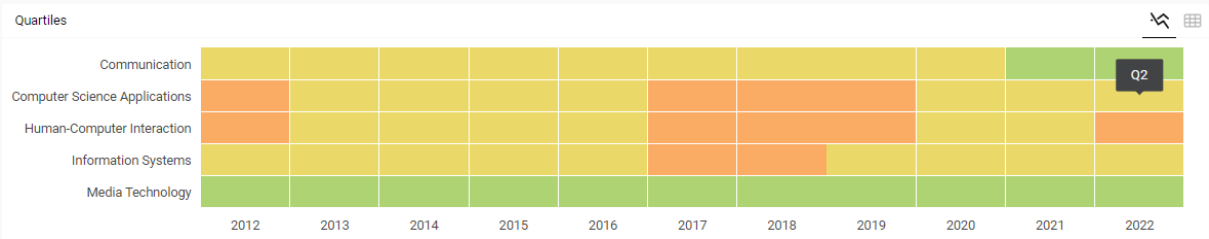
SJR Scimago Journal & Country Rank

Enter Journal Title, ISSN or Publisher Name

Home Journal Rankings Country Rankings Viz Tools Help About Us

Social Network Analysis and Mining

COUNTRY	SUBJECT AREA AND CATEGORY	PUBLISHER	H-INDEX
Austria Universities and research institutions in Austria Media Ranking in Austria	Computer Science Computer Science Applications Human-Computer Interaction Information Systems Engineering Media Technology Social Sciences Communication	Springer-Verlag Wien	43



2. ACM Transactions on Database Systems

Article [J2] : A Mechanism for Generating OLAP Queries Dynamically based on User Preferences and Hierarchical Visualization Constraint.

Auteurs : **Hamdi Ghassen,** Telli Abdelmoutia and Omri Mohamed Nazih

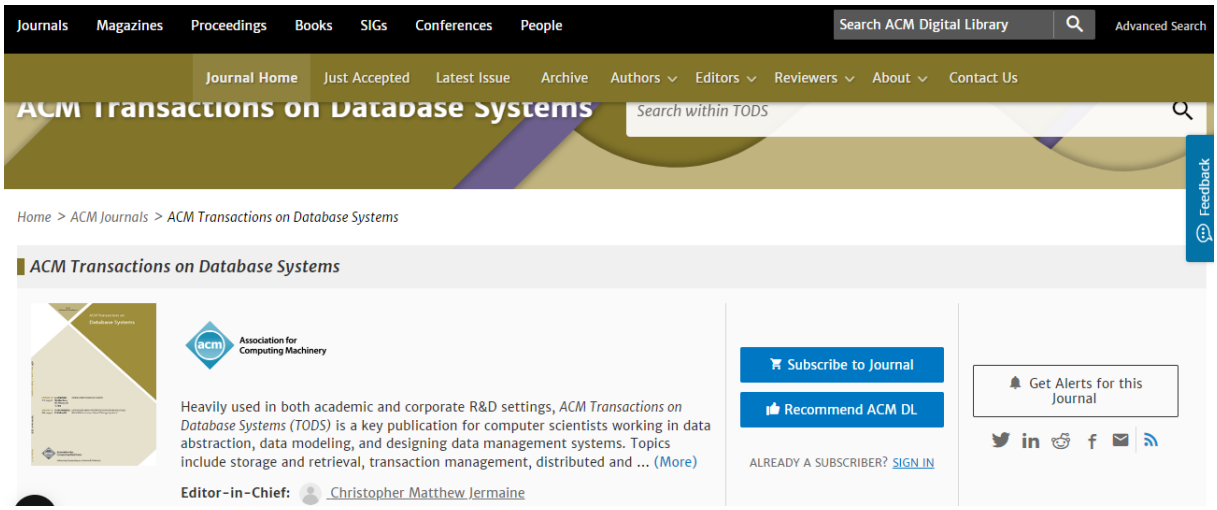
Journal : ACM Transactions on Database Systems

Etat : Under Review

Quartile : Q1

SJR : 1.71

ISSN: 0362-5915


Journals
Magazines
Proceedings
Books
SIGs
Conferences
People
Search ACM Digital Library
Advanced Search
Journal Home
Just Accepted
Latest Issue
Archive
Authors
Editors
Reviewers
About
Contact Us
ACM Transactions on Database Systems
Search within TODS
Feedback
Home > ACM Journals > ACM Transactions on Database Systems
ACM Transactions on Database Systems
Association for Computing Machinery
Heavily used in both academic and corporate R&D settings, ACM Transactions on Database Systems (TODS) is a key publication for computer scientists working in data abstraction, data modeling, and designing data management systems. Topics include storage and retrieval, transaction management, distributed and ... (More)
Editor-in-Chief: Christopher Matthew Jermaine
Subscribe to Journal
Recommend ACM DL
ALREADY A SUBSCRIBER? SIGN IN
Get Alerts for this Journal

ScholarOne Manuscripts™ | Ghassen Hamdi | Instructions & Forms | Help | Log Out

Transactions on Database Systems

Home | Author | Review

Author Dashboard

Author Dashboard

1 Submitted Manuscripts

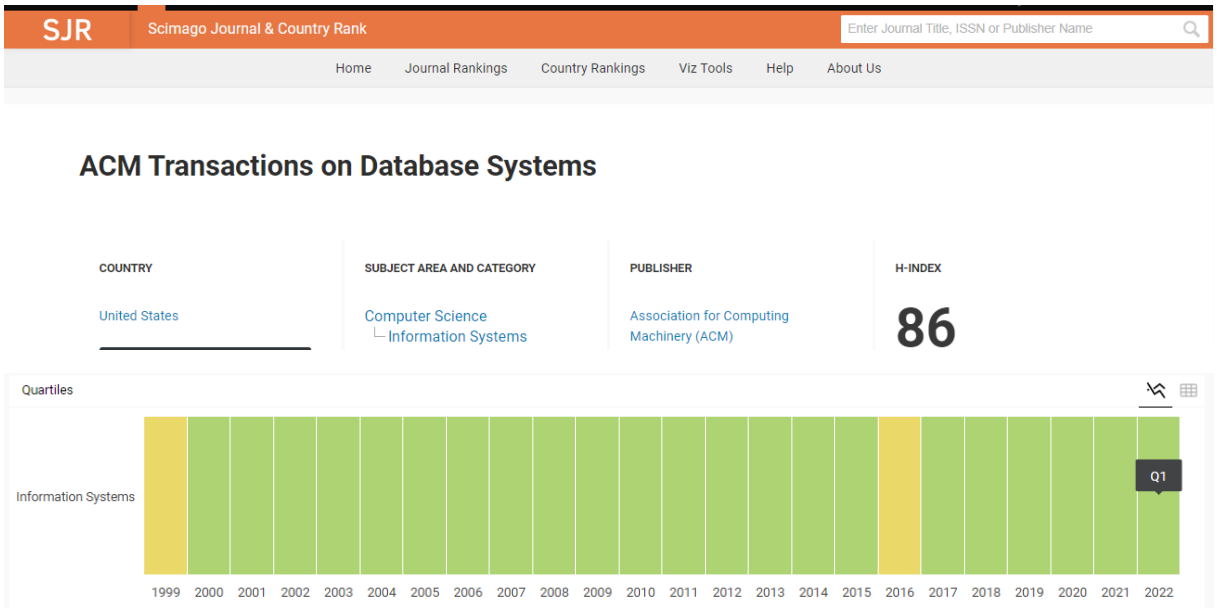
Start New Submission

Legacy Instructions

5 Most Recent E-mails

Submitted Manuscripts

STATUS	ID	TITLE	CREATED	SUBMITTED
Under Review	TODS-2023-0072	A Mechanism for Generating OLAP Queries Dynamically based on User Preferences and Hierarchical Visualization Constraint View Submission Cover Letter	22-Nov-2023	23-Nov-2023


SJR
Scimago Journal & Country Rank
Enter Journal Title, ISSN or Publisher Name
Home
Journal Rankings
Country Rankings
Viz Tools
Help
About Us
ACM Transactions on Database Systems
COUNTRY
United States
SUBJECT AREA AND CATEGORY
Computer Science
Information Systems
PUBLISHER
Association for Computing Machinery (ACM)
H-INDEX
86
Quartiles
Information Systems
Q1
1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022

ACM Transactions on Database Systems
Q1
Information Systems
best quartile
SJR 2022
1.71
powered by scimagojr.com

3. Computational and Applied Mathematics

Article [J3] : Enhancing DL-Lite Possibilistic Logic : Interval-Based Knowledge Representation and Reasoning with Degrees of Inconsistency.

Auteurs : **Hamdi Ghassen,** Telli Abdelmoutia and Omri Mohamed Nazih

Journal : Computational and Applied Mathematics

Etat : Under Review

Quartile : Q2

SJR : 0.58

ISSN: 1807-0302

Overview

Computational & Applied Mathematics began in 1981, as the main scientific publication of the Brazilian Society of Computational and Applied Mathematics (SBMAC).

The objective of the journal is the publication of original research in Applied and Computational Mathematics, with interfaces in Physics, Engineering, Chemistry, Biology, Operations Research, Statistics, Social Sciences and Economy. The journal has the usual quality standards of scientific international journals, and we aim high level of contributions in terms of originality, depth and relevance. CAM is currently reviewed in Mathematical Reviews and Institute of Scientific Information (Webofscience).

Editor-in-Chief
Jose Eduardo Souza de Cursi

Impact factor
2.6 (2022)

5 year impact factor
2.2 (2022)

Submission to first decision (median)
18 days

Downloads
162,789 (2022)

Computational and Applied Mathematics

Ghassen Hamdi | Logout

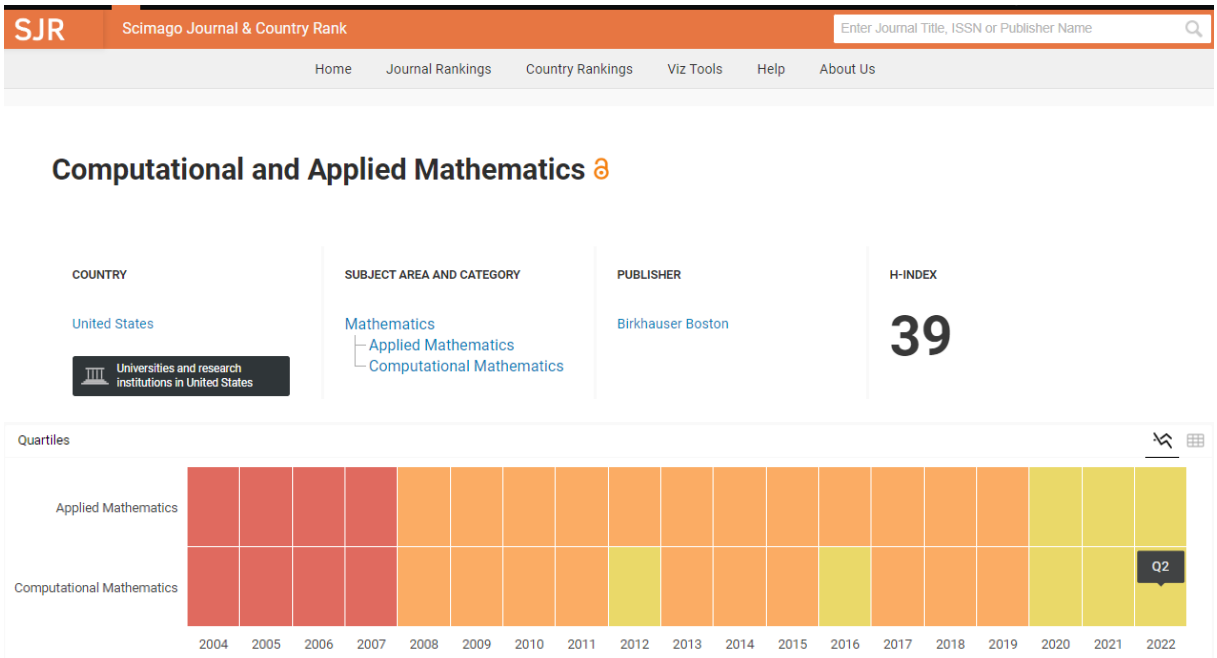
Home Main Menu Submit a Manuscript About Help

Submissions Being Processed for Author

Page: 1 of 1 (1 total submissions)

Results per page 10

Action	Manuscript Number	Title	Initial Date Submitted	Status Date	Current Status
Action Links	COAM-D-24-00332	Enhancing DL-Lite Possibilistic Logic Interval-Based Knowledge Representation and Reasoning with Degrees of Inconsistency	13 Feb 2024	16 Feb 2024	Editor Assigned


SJR
Scimago Journal & Country Rank
Enter Journal Title, ISSN or Publisher Name
Home
Journal Rankings
Country Rankings
Viz Tools
Help
About Us
Computational and Applied Mathematics
COUNTRY
United States
Universities and research institutions in United States
SUBJECT AREA AND CATEGORY
Mathematics
Applied Mathematics
Computational Mathematics
PUBLISHER
Birkhauser Boston
H-INDEX
39
Quartiles
Applied Mathematics
Computational Mathematics
Q2
2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022

Computational and Applied Mathematics
Q2
Applied Mathematics
best quartile
SJR 2022
0.58
powered by scimagojr.com

4. Computer Science and Information Systems

Article [J4] : Incremental and Hierarchical Visualization of Query Results on a Data Warehouse.

Auteurs : **Hamdi Ghassen**

Journal : Computer Science and Information Systems

Etat : Under Review

Quartile : Q3

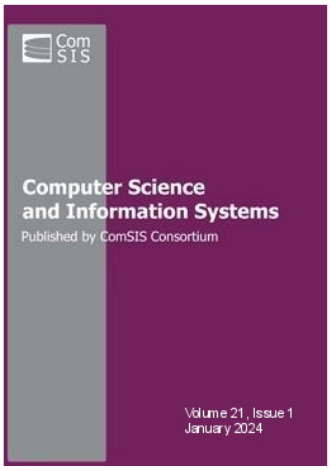
SJR : 0.32

ISSN: 2406-1018

Computer Science and Information Systems

Computer Science and Information Systems

The international journal published by ComSIS Consortium



ComSIS publishes original papers of lasting value covering both theoretical foundations of computer science and commercial, industrial, or educational aspects that provide new insights into design and implementation of software and information systems. ComSIS also welcomes surveys papers that contribute to the understanding of emerging and important fields of computer science. In addition to wide-scope regular issues, ComSIS also includes special issues covering specific topics in all areas of computer science and information systems.

NEWS: Recently introduced publication processing fee for papers submitted to ComSIS Journal. (More information)

About the journal
- Home page
- Contact information
- Aims and scope
- Indexing information
- Editorial policies
- ComSIS consortium
- Journal boards
- Managing board

For authors
- Information for contributors
- Paper submission
- Article submission through OJS
- Copyright transfer form
- Download section

For readers
- Forthcoming articles

Computer Science and Information Systems

HOME ABOUT USER HOME

Home > User > Author > **Active Submissions**

Active Submissions

ACTIVE ARCHIVE

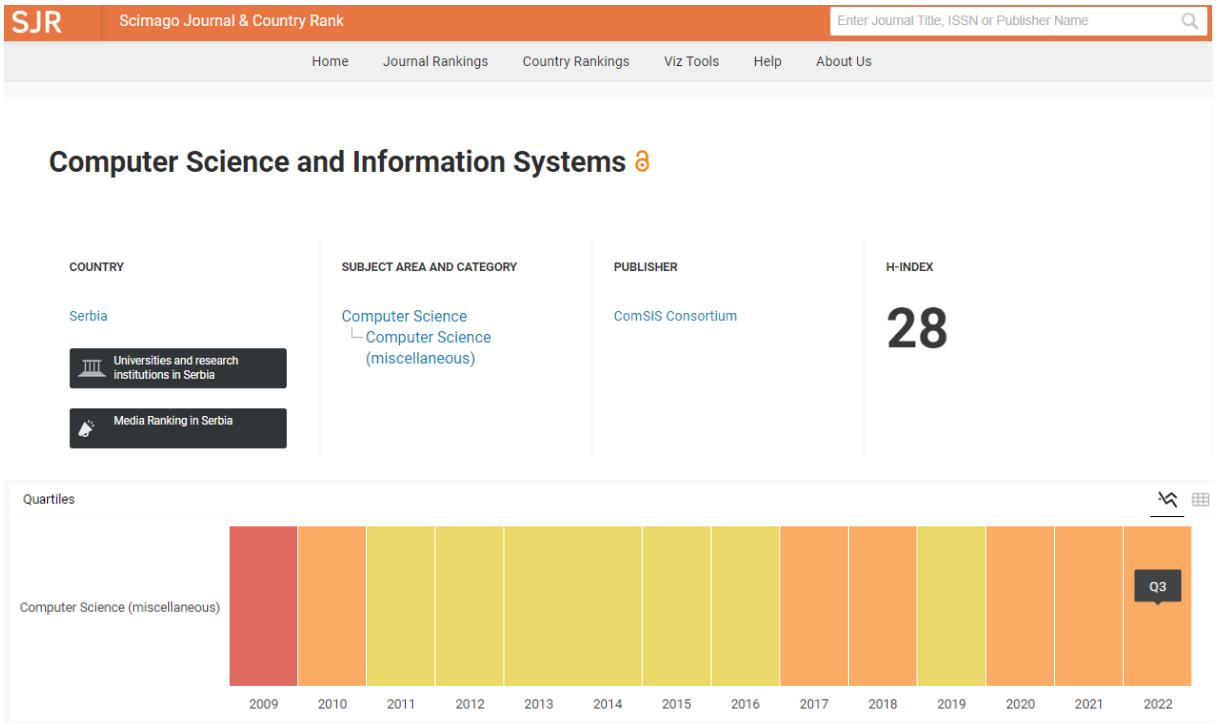
ID	MM-DD SUBMIT	SEC	AUTHORS	TITLE	STATUS
16984	02-01	ART	Hamdi	INCREMENTAL AND HIERARCHICAL VISUALIZATION OF QUERY...	Awaiting assignment

1 - 1 of 1 Items

Start a New Submission

CLICK HERE to go to step one of the five-step submission process.

ISSN: 2406-1018


SJR
Scimago Journal & Country Rank
Enter Journal Title, ISSN or Publisher Name
Home
Journal Rankings
Country Rankings
Viz Tools
Help
About Us
Computer Science and Information Systems
COUNTRY
Serbia
Universities and research institutions in Serbia
Media Ranking in Serbia
SUBJECT AREA AND CATEGORY
Computer Science
Computer Science (miscellaneous)
PUBLISHER
ComSIS Consortium
H-INDEX
28
Quartiles
Computer Science (miscellaneous)
Q3
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022

Computer Science and Information Systems
Q3
Computer Science (miscellaneous)
best quartile
SJR 2022
0.32
powered by scimagojr.com

III. Articles publiés dans des Conférences Internationales

1. International Conference on Cyberworlds (CW 2023)

Article [C1] : A Personalized Multidimensional Navigation in a Limited Visualization Context

Auteurs : **Hamdi Ghassen** and Omri Mohamed Nazih

Conférence : International Conference on Cyberworlds (**CW** 2023), 3-5 October 2023. Sousse, Tunisia, pp. 54--61

Classement : **C**

A Personalized Multidimensional Navigation in a Limited Visualization Context

Publisher: IEEE

Cite This PDF

Ghassen Hamdi ; Mohamed Nazih Omri All Authors

9
Full
Text Views

Abstract

Document Sections

I. Introduction

II. Related Works

III. Motivation

IV. Data Warehouse Personalization

V. Presentation of the PersonalizationDW Tool

Show Full Outline

Authors

Figures

References

Abstract:
The needs of decision-makers vary greatly depending on their specialties, expertise and knowledge and therefore they do not visualize decision-making data in the same way. Everyone can have their preferences and prerequisites. The multidimensional model, made up of a set of facts analyzed according to different dimensions, is intended for a set of decision-makers and includes all the indicators and axes of analysis of a particular domain. They do not consider the preferences of each decision maker. A personalization phase of this diagram according to the preferences of these decision-makers will improve the decision-making process. In addition, the limited capacity of the devices used (smartphones, tablets, etc.) for consulting the results of OLAP queries presents another challenge for users. In this context, we propose to express the preferences in the form of visualization constraints which will be applied to the schema to represent the result in an incremental way.

Published in: 2023 International Conference on Cyberworlds (CW)

Date of Conference: 03-05 October 2023

DOI: 10.1109/CW58918.2023.00018

Date Added to IEEE *Xplore*: 06 December 2023

Publisher: IEEE

ISBN Information:

Conference Location: Sousse, Tunisia

ISSN Information:

11/02/2024 22:36 decision on your paper 4962 - hamdi.ghassan@gmail.com - Gmail

Gmail cw

99+
Mail
1
Chat
Meet

Nouveau message

Boîte de récep... 2080
Messages suivis
En attente
Important
Messages envoyés
Brouillons 120
Catégories
Plus

Libellés

Personnel
Unwanted
Plus

Traduire en français

Dear Ghassen Hamdi,

We are very happy to inform you that your paper 4962: "A personalized multidimensional navigation in a limited visuali

is ACCEPTED as a FULL PAPER by the 2023 International Conference on Cyberworlds (CW2023).

Please find the comments from the reviewers attached and make all the necessary amendments for a camera-ready ve

Your paper will be published in the conference proceedings printed by IEEE Conference Publishing Services (CPS), an
Xplore Digital Library, IEEE Computer Society Digital Library and reference databases of all major referencing indices.
issues with extended versions of the best papers. This will be decided after the conference.

Please prepare a camera-ready version of your paper which has to be written in English, carefully proofread, and forma
A4. The page limit is 8 pages. The camera-ready paper must be submitted through the Author Submission Site by 28 Ju

https://ieeecps.org/#!/auth/login?ak=1&pid=4TuWFPhXEpm2eDTBx6HP5j

The Online Author Kit has a PDF conversion and verification tool. The publisher will check your paper for plagiarism.

For each accepted full paper a registration fee has to be paid by 28 July. Refer to the conference web page https://cw2

The details of the online option of the presentation will be emailed to the authors in September.

BY REPLYING TO THIS EMAIL PLEASE CONFIRM THAT:
/1/ The paper will be registered by 28 July.
/2/ Indicate the planned attendance: online or real.
/3/ Give the name and contact email of the presenting author.

CW Search by: All Source: CORE2023

Search

Showing results 1 - 8 of 8

Export

CORE2023 Summary:
A* - 7.47% of 803 ranked venues
A - 14.45% of 803 ranked venues
B - 27.77% of 803 ranked venues
Australasian B - 0.75% of 803 ranked venues
C - 47.2% of 803 ranked venues
Australasian C - 2.37% of 803 ranked venues
Other - 164 total

Title	Acronym	Source	Rank	Note	DBLP	Primary FoR	Comments	Average Rating
ACM Conference on Computer Supported Cooperative Work	CSCW	CORE2023	A	none	view	4608	1	5.0
European Conference on Computer Supported Cooperative Work	ECSCW	CORE2023	B	none	view	4608	0	N/A
Heterogeneity in Computing Workshop	HCW	CORE2023	C	See note	none	4606	0	N/A
IEEE International Conference on Web Services	ICWS	CORE2023	A	none	view	4606	0	N/A
International Conference on Computer Supported Cooperative Work in Design	CSCWD	CORE2023	C	See note	view	4608	0	N/A
International Conference on Cyberworlds (was International Symposium on Cyberworlds)	CW	CORE2023	C	See note	view	4607	2	5.0
International Conference on Web and Social Media	ICWSM	CORE2023	A	none	view	4601	2	3.0
International Conference on Web Engineering	ICWE	CORE2023	B	none	view	4605	4	4.5

A personalized multidimensional navigation in a limited visualization context

Ghassen Hamdi
MARS Research Laboratory,
University of Sousse, Tunisia
hamdighassan@gmail.com

Mohamed Nazih Omri
MARS Research Laboratory,
University of Sousse, Tunisia
mohamednazih.omri@eniso.u-sousse.tn

2023 International Conference on Cyberworlds (CW) | 979-8-3503-1565-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/CW58918.2023.00018

***Abstract*—The needs of decision-makers vary greatly depending on their specialties, expertise and knowledge and therefore they do not visualize decision-making data in the same way. Everyone can have their preferences and prerequisites. The multidimensional model, made up of a set of facts analyzed according to different dimensions, is intended for a set of decision-makers and includes all the indicators and axes of analysis of a particular domain. They do not consider the preferences of each decision maker. A personalization phase of this diagram according to the preferences of these decision-makers will improve the decision-making process. In addition, the limited capacity of the devices used (smartphones, tablets, etc.) for consulting the results of OLAP queries presents another challenge for users. In this context, we propose to express the preferences in the form of visualization constraints which will be applied to the schema to represent the result in an incremental way.**

***Index Terms*—Data warehouse, OLAP queries, hierarchical visualization under constraints, Personalization**

I. Introduction

A. Context and issues

The needs of decision-makers vary greatly depending on their specialties, expertise and knowledge and therefore they do not visualize decision-making data in the same way. Everyone can have their preferences and prerequisites. The multidimensional model made up of a set of facts analyzed according to different dimensions is intended for a set of decision-makers and includes all the indicators and axes of analysis of a particular domain. They do not consider the preferences of each decision maker. A personalization phase of this diagram according to the preferences of these decision-makers will improve the decision-making process. In addition, the limited capacity of the devices used (smartphones, tablets, etc.) for consulting the results of OLAP queries presents another challenge for users. In this context, we propose to express the preferences in the form of visualization constraints which will be applied on the diagram to represent the result in an incremental way. In this article, we first formalize the ”User profile” concept dedicated to data warehouses. Also, we detail our personalization method based on the incremental representation of the result to the user. This visualization takes into consideration the display space constraint. In data warehouse personalization works, the concept of ”hierarchy” that characterizes the data structure in the data warehouse, is ignored by quite a few personalization works ([1], [2]). In these works the personalization of the results of the analyzes carried out on the data, has become a priority which neglects the hierarchical aspect under which multidimensional data returned to the user are stored in the warehouse. However, the limited capacity of the devices used to consult data from a data warehouse is an obstacle for the user who wishes to view all the data, whereas the solutions proposed in the literature ([1]) are limited to returning the most important preferences while respecting the constraint. Thus, other preferences, which are also important, are excluded from the result of the user's query.

B. Contributions

In this article, we propose to personalize OLAP queries by visualizing data according to the different levels of detail preferred by users. Indeed, the users of a data warehouse request results with different degrees of refinement for certain axes of analysis, in order to have a more or less detailed global vision (according to the levels of refinement requested). However, this personalized hierarchical view of multidimensional data offered to users is not unlimited (free). It is limited by the data visualization constraint related to the capacity of the device used to display the results. We provide the user with a hierarchical view of the results, based on his preferences expressed in his profile. In addition, in order to weaken the effect of the visualization constraint on the personalized results, we offer the user an incremental visualization of the data returned by the personalized query, in a limited context. Our method of personalizing decisional queries is not restricted to the visualization of dimensions with a single attribute ([1]), although we are placed in a context under visualization constraint. Our method guarantees users a hierarchical and incremental view of the data in a restricted visualization environment, taking into account the preferences expressed not only on the structures of the multidimensional model ([3], [4]) but also on data from the warehouse. To do this, we propose a profile model that allows to express the preferences on the multidimensional model and the warehouse data. Our profile model also allows to exploit the hierarchical characteristic in order to interpret new preferences not defi ned by the user. In this work, we extend OLAP queries expressed with the MDX (Multi-Dimensional eXpressions) language developed by Microsoft.

979-8-3503-1565-3/23/$31.00 ©2023 IEEE

DOI 10.1109/CW58918.2023.00018

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:01:51 UTC from IEEE Xplore. Restrictions apply.

The remainder of this article is arranged as follow. In section II we present the main approaches that tried to solve the problem of personalization in the litterature. Section III will be dedicated to the presentation of the motivation of our personalization method. Section IV details our proposed method of data warehouse personalization which allows to personalize the multidimensional structures and the data of the warehouse by adopting the quantitative approach in order to define the preferences of the users. Section V introduces the PersonalizationDW tool which performs a personalized OLAP query under constraint. Section VI provides the experimental study conducted in this work and analyse the obtained results. In the final section VII, we present a summary of the work as well as some ideas for future projects.

II. Related Works

Many personalization works are offered in the field of databases, whose personalization methods followed do not treat the problem of personalization in the same way. This work can be classified according to two main approaches: the quantitative approach, which consists in expressing the preferences indirectly by the user by associating numerical scores calculated by score functions, and the qualitative approach where the preferences are specified directly by binary relations.

The personalization method proposed by [5] presents the preferences of the user profile defined by score functions and operated on the content. This method does not support constraints. In fact, the user profile only includes preferences which establish a total order on the tuples, and which are used as filters on the answers of the request. The tuples are compared after the query result is calculated and only those with significant interests will be displayed to the user. This method displays the best tuples of the response to a query according to the preferences of the profile.

In [6], the authors suggest a personalization method based on the preferences of the user profile expressed on the content. This method defined in a quantitative framework makes it possible to select the most relevant preferences for the user. These preferences are represented by score functions which assign degrees of interest to atomic elements of the query, for example, a score must be assigned to a selection predicate or a join predicate. These preferences then establish a total order on the queries: one query is more interesting than another if its degree of interest is higher. This method thus consists in enriching the user request with the preferences of the user profile in order to maximize the interest of the user in the response to the request. Indeed, to personalize a given query, the algorithm determines the top K preferences that are not contradictory with the query.

In [7], [8], the authors offers a personalization method that does not present constraints. Profile preferences are set on content. Creating binary relationships between query response tuples. These preferences are represented by a binary relation or a strict partial order. They are introduced in the relational language, in [8], by the preference operator WINNOW. Thus, a query with preferences is a relational algebra query containing at least one occurrence of the WINNOW operator. In addition, [8] proposes properties to optimize the calculation of preferred tuples from the calculated result of the query, ie after accessing the data.

In the personalization method proposed by [9], the preferences of the user profile relate to the content and define an ordering relationship between the tuples of the response to the query. The objective of this method is to return the most interesting tuples of the result based on user preferences, and which satisfy a given condition. To solve this same problem, [10] takes into account constraints, called hard constraints, which are conditions that must be satisfied.

In [11], the authors propose their method of personalization where the preferences of the user profile relate to the form (or the presentation) and define an order relation between the tuples of the response to the query. This method does not consider constraints. In this article, two tuples are compared after accessing data sources. The goal is to find the most interesting tuples with respect to user preferences, and which correspond to a given selection of attributes. The presentation is expressed by the selection of relevant attributes and by displaying the detail of the tuples which correspond to them in tables and in an order according to their importance from the point of view of the user. Contrary to the ordering of tuples on display, the selection of attributes is done before accessing the data. Furthermore, the calculation of the scores of a tuple depends on a set of specified attributes indicating which attributes should be used in this calculation. Consequently, several different scores can be associated with the same tuple. This problem is solved by transforming it into a problem of choosing the most useful set of attributes, i.e. the most influential Top-m attributes in the computation of the Top-n ordered tuples of the response, with $m << M$ attributes and $n << N$ tuples.

In their work, [2], [12], [13], [27] propose a personalization approach which consists in personalizing the schema of a data warehouse by creating new levels of hierarchy. The objective of this work is to be able to exploit and share decision makers' knowledge stored in dimensions as aggregations of new hierarchies in the form of "if-then" type rules, during the warehouse query process.

In [3], [26], the authors propose a context-aware preference model that is defined on a multidimensional database schema. Using a qualitative approach, OLAP preferences are modeled and closely depend on the analysis context of the user. This is why a conceptual model of the OLAP context is designed using a tree of OLAP analysis elements. There are preferences between dimensions and others concerning the same dimension.

In [14], the authors extend Favre's proposal by proposing to create the new level of hierarchy after the validation of this appropriate adaptation by the user.

In [15], the authors propose an approach based on the idea of creating materialized views of a data warehouse according to each user profile. This profile is composed by a set of

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:01:51 UTC from IEEE Xplore. Restrictions apply.

permanent preferences called rigid preferences and temporary preferences called flexible preferences.

In [16]–[18], the authors propose to personalize the data warehouse schema at the conceptual level. The personalized OLAP schema is tailored to the specific characteristics of a decision maker.

In [1], [19], the authors propose an approach for the personalization of OLAP query answers and provide an adapted visualization for a user based on his profile. The latter contains the preferences that classify the values of all the dimensions and the visualization constraints that control the form of visualization of the results of the query.

In order to personalize the analyses, a personalization approach based on the construction of preferences is proposed by [20], [21], [25]. The proposed conceptual model is based on multidimensional concepts (fact, dimension, hierarchy, measures, parameter or attribute). The contents of the multidimensional tables are adapted based on a set of user preferences which are expressed in the form of a priority order on the attributes of the multidimensional model.

In their work, [22], [23] propose to adapt the expression of MDX queries according to the preferences of the user. This approach allows the expression of preferences on the OLAP query. An OLAP algebra (the MYOLAP algebra), which allows users to declare their preferences on attributes, measures and the level of aggregation is proposed.

In [24], the authors extend Golfarelli and al's approach by applying association rules in the MDX query log to implicitly extract user preferences and annotate their queries with MYOLAP algebra.

III. Motivation

We take the example of the query below where the user queries the Sales cube, the schema of which is shown in the figure 1. This request expresses the user's need to see the measures (NumberSale) detailed according to three levels for the Customers dimension (Country, State Province, City) and four levels for the Product dimension (Family, Department, Category, Subcategory).

$$"SELECT\ \{[customer].[hcustomer].[country],$$
$$[customer].[hcustomer].[stateprovince],$$
$$[customer].[hcustomer].[city]\}ONCOLUMNS,$$
$$\{[product].[HProduct].[family].members,$$
$$[product].[HProduct].[category].members,$$
$$[product].[HProduct].[subcategory].members\}ONROWS$$
$$FROM[SalesCube]where([Measures].[unitsales])"$$

This query is used to return data, with different levels of detail for the same dimension, viewed in a cross tab. (See Table I).

Nevertheless, the size of the result returned by this query is very high, which makes the user's data analysis task more complex, especially if he has a device with a low viewing capacity (exp: PDA, iPhone, etc.). To do this, we propose to personalize the query in order to offer the user the most relevant data according to the different preferred levels of detail while respecting the visualization constraint.

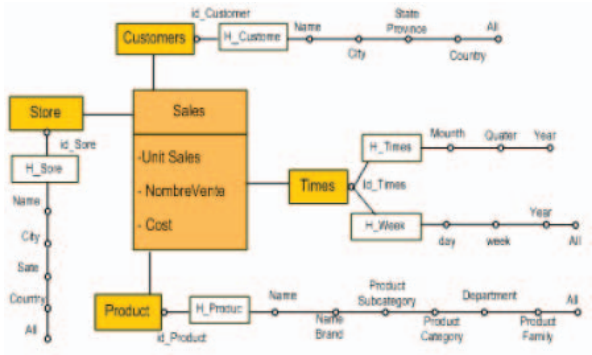


Fig. 1. schema of the Sales cube

Unit Sales		Country		State province		...
		Algeria	Tunisia	Aneba	...	...
Family	Drink	331	442	32	...	...
	Food	125	23	56	...	...
Category	Hot Beverages	203	67	27	...	...
	Bread	286	131	64	...	...
	Fruit	27	10	29	...	...
Subcategory	Coffee	38	83	49	...	...
	Bagel	98	13	26	...	...
	...	...	...	...	...	...

TABLE I
Query result.

We offer the user a personalized multidimensional navigation through the cube data, in a limited visualization context. This navigation allows you to zoom in on the one hand on the most preferred analysis parameters of a dimension, and on the other hand on the preferred members which are the instances of these parameters and the references of the measures in the cube. In addition, this browsing requires in certain cases to extract, from a preferred level of analysis, members without knowing their degree of preferences. Our profile model makes it possible to cover this lack of definition of preferences by the user, and to interpret new preferences in order to complete the definition of the decisional profile. The deduction of new preferences, in our profile model, is mainly based on the hierarchical relationship between the members of a dimension. However, a multidimensional navigation, in a restricted viewing context, does not make it possible to satisfy the user's need to see all his preferences. To meet this need, our personalization method makes it possible to return to the user an incremental personalized result respecting the preferences expressed and the constraint imposed.

Example 1: A user who prefers to analyze the results of two dimensions Product and Customer, respectively according to the parameters City, State Province and Country following the dimension Customer and Family, Subcategory and Category following the dimension Product. In addition, the user favors the data of the "Food" product family, of its "Fruit" category, of the "Bagels" and "Fresh Fruit" subcategories, and of the

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:01:51 UTC from IEEE Xplore. Restrictions apply.

”Drink” family, of its ”Hot Beverages” category and its subcategory ”Coffee”.
We assume that the device used by the user can only display three analysis parameters horizontally and two vertically. At member level, the device only displays a maximum of three columns and seven rows.

Unit Sales		Country		State province
		Algeria	Tunisia	Wahran
Family	Food	331	442	34
	Drink	125	110	26
Category	Bread	255	120	36
	Fruit	76	362	27
	Hot Beverages	126	110	12
Subcategory	Bagel	256	120	32
	Fresh Fruit	47	48	98

TABLE II
INCREMENTAL PERSONALIZED RESULT (THE FIRST ITERATION)

Unit Sales		State Province		
		Aneba	Sidi Bouzid	Sfax
Subcategory	Coffee	125	112	110

TABLE III
INCREMENTAL PERSONALIZED RESULT (THE SECOND ITERATION)

To display the results of tables II and III, we propose to determine, first, the analysis parameters used in the query and preferred by the user, taking into account the visualization constraint. Then, we determine, incrementally, the instance members of the selected parameters, respecting the visualization constraint. Finally, for each iteration, we rewrite the query according to the determined preferences. In our example, we get the personalized query rewritten as follows, for iteration 1:

”$SELECT\{[customer].[hcustomer].[country].[Algeria],$
$[customer].[hcustomer].[country].[Tunisia],$
$[customer].[hcustomer].[stateprovince].[wahran]$
$\}ONCOLUMNS,$
$\{[product].[HProduct].[family].[Food],$
$[product].[HProduct].[family].[Drink],$
$[product].[HProduct].[category].[Bread],$
$[product].[HProduct].[category].[Fruit],$
$[product].[HProduct].[category].[HotBeverages],$
$[product].[HProduct].[subcategory].[Bagel],$
$[product].[HProduct].[subcategory].[FreshFruit]$
} ON ROWS
FROM [SalesCube] where ([Measures].[unitsales])”

IV. DATA WAREHOUSE PERSONALIZATION

Personalized data warehouses form a context characterized by the union of the multidimensional aspects of data and user profiles. A personalized data warehouse Ep is described by a multidimensional conceptual model, profiles and preference expression functions:
$Ep = (Sch(C), P^C, F^C)$ with: Sch(C) : represents the multidimensional schema of the data warehouse: the constellation schema.
We use the constellation model which groups together several subjects of analysis (facts) studied according to different axes of analysis (dimensions) which may be shared. These dimensions are composed of several hierarchies.
$P^C = \{P_1, ..., P_u\}$: represents the list of warehouse user profiles.
$F^C = \{f_1, ..., f_z\}$: is the set of functions for personalizing the data warehouse and setting user preferences.
In this work, we propose to personalize the multidimensional structures and the data of the warehouse by adopting the quantitative approach in order to define the preferences of the users. We use scoring functions that allow users to assign importance weights between 0 and 1 reflecting the degree of interest in that structure or data.

A. The representation of the profile

In the context of our work, the profile consists of viewing preferences and constraints. Preferences are expressed on multidimensional structures (measures, dimensions, hierarchies, etc.) and on data (instances). The visualization constraint, which is linked to the device used by the user to interrogate the data cube, delimits the number of result data delivered.
A profile P can be defined by: $P = (Pr_S, Pr_D, C_v)$ with:

- Pr_S is the set of preferences expressed on multidimensional structures.
- Pr_D is the set of preferences expressed on the instances.
- C_v represents the visualization constraint characterizing the context of visualization of the results for the user.

1) The formalization of preferences: Preferences are defined using functions of F which are essentially based on score functions which assign a value between 0 and 1 representing the degree of interest of the decision maker. The preferences that make up a profile are expressed on the multidimensional structures and on the data.

2) Structure Preferences: The user can define his preferences on the multidimensional structures he wishes to visualize in the results of the queries, by assigning degrees of importance. Let $Weight_z \in F$, a function called : $Weight_z : Z \rightarrow [0, 1]$, which assigns a weight between 0 and 1 to a multidimensional structure $z \in Z, Z \subseteq Sch(C)$.
Preferences are expressed on dimensions, hierarchies, measures, aggregation functions, etc. by assigning weights between 0 and 1 using the function $Weight_z$.
Consider the example of setting preferences on dimensions, using the function $Weight_D$.
$Weight_D : D \rightarrow [0, 1] \ \forall i \in [1, n] \, and \, D_i \in D, Weight_D(D_i) \in [0, 1]$

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:01:51 UTC from IEEE Xplore. Restrictions apply.

Example 2: $D = \{Customers, Product, Times, Store\}$. $Weight_D(Customers) = 0.65; Weight_D(Product) = 0.9; Weight_D(Times) = 0.85; Weight_D(Store) = 0.75$
The function $Weight_D(D_i)$ provides a general weight on a dimension without specifying the weight of these sub-elements (hierarchies, levels). The final weight that can be attributed to a dimension must take these elements into account, for this we define the notion of intra-dimension weight.
A second weight is described to define an order relation between the elements of different dimensions called extra dimension weight. Thus, we distinguish two types of weights attributed to the OLAP preferences defined on the analysis axes: an intra-dimensional weight, noted by PI, and an extra-dimensional weight, noted by PE.

a) Intra dimension weight: An intra-dimension weight is assigned by the user to the hierarchies and analysis parameters of a dimension. This score only compares between hierarchies or parameters within the same dimension.
The Weight function defined previously is used to declare these weights.

- The intra-dimension weight of a hierarchy: The user can give importance weights to the hierarchies of a dimension $D_i \in D$ using the function : PI^{Di} :

  $$H^{Di} \rightarrow [0,1]\ \forall H_j^{Di} \in H^{Di}, PI_{H^{Di}}(H^{Di}j) \in [0,1].$$

  Example 3: Let's take the Times dimension:
  $PI_{H^{Times}}(H_Times) = 0.9$; $PI_{H^{Times}}(H_week) = 0.3$
- The intra-dimension weight of a parameter: The user can define his preferences on the parameters with which he prefers to see the results. This using the function:

  $$PI_{A^{Di}} : A^{Di} \rightarrow [0,1]\ \forall a \in A^{Di}, PI_{A^{Di}}(a) \in [0,1].$$

Example 4: We take the Times dimension as an example: $A^{Times} = \{PI_{A^{Times}}(idTimes) = 0.1; PI_{A^{Times}}(Day) = 0.2; PI_{A^{Times}}(Mounth) = 0.75; PI_{A^{Times}}(week) = 0.85; PI_{A^{Times}}(Year) = 0.95; PI_{A^{Times}}(all) = 0.5\}$

b) Extra dimension weight: An extra dimension weight, denoted by PE, is calculated based on the PI and the weight of the dimension. It is used to indicate the importance of hierarchies and parameters of different dimensions to compare them with each other.

- The extra dimension weight of a hierarchy: The PE of a hierarchy is calculated according to its PI and the weight of the dimension to which it belongs. $\forall i \in [1,n]\, and\, D_i \in D, \forall H_j^{Di} \in H^{Di} : PE(H_j^{Di}) = PI(H_j^{Di}) * Weight_D(Di)$.

Example 5: We take the hierarchies from the Times dimension.
$PE(H_Times) = PI_{H^{Times}}(H_Times) * Weight_D(Times) = 0.9 * 0.85 = 0.765;$
$PE(H_week) = 0.255$

- The extra dimension weight of a parameter: The PE of an analysis parameter depends not only on its PI and the weight of the dimension, but also on the maximum PI of the hierarchies to which it belongs; since a parameter can belong to one or more hierarchies in a dimension. $\forall a \in A^{Di}, let H_a = \{H_1, ..., H_b\}\, such\, as\, \forall j \in [1,b], a \in H_j \wedge H_j \in H^{Di} : PE(a) = PI_{A^{Di}}(a) * max(\{PI_{H^{Di}}(H_j)\ j \in [1,b]\}) * Weight_D(D_i) = PI_{A^{Di}}(a) * max(\{PE(H_j)\ j \in [1,b]\})$.

Example 6: $PE(Year) = PI_{A^{Time}}(Year) * max(\{PE(H_week), PE(H_Times)\}) = 0.95 * 0.765 = 0,727$.

3) Data preferences: In a warehouse, data is stored as multidimensional tuples representing dimension members and measures in the fact table. A multidimensional data tuple is an instantiation of multidimensional structures (dimension parameters and measures). Expressing preferences on warehouse data helps identify preferred tuples for the user. The scoring functions allow you to assign weights between 0 and 1, indicating the interest of the user, to the values of the domains of parameters and measures.
To express preferences on the values of a measure or a parameter, we have:
$Op = \{" = ", " < ", " > "\}\, and\, Weight_{parameter} : A^{Di}\, x\, Op\, x\, dom(A^{Di}) \rightarrow [0,1]\ such\, as\ \forall i \in [1,n], D_i \in D\, and\, \forall a \in A^{Di}, \forall v \in dom(a)\, and\, op \in Op,$
$Weight_{parameter}(a, op, v) \in [0,1]$.

Dimension	Parameter	Value	Weight
Product	Family	Food	0.95
Product	Family	Drink	0.85
Product	Subcategory	Bagel	0.9
Product	Subcategory	Coffee	0.45
Product	Category	Hot Beverages	0.7
Product	State province	Sousse	0.9
Product	State province	Mahdia	0.82
Product	City	Enfidha	0.95
Product	City	Cheba	0.85

TABLE IV
EXAMPLE OF DATA PREFERENCES

4) Interpreted preferences: Our method of expressing preferences makes it possible to infer new preferences on the data, not expressed by the user. Indeed, we can take advantage of the hierarchical structure of multidimensional data in order to interpret preferences that are not defined directly by the user. Indeed if the decision maker has a preference for the city of "Tunis" then we can infer that he has a certain preference for the father member "Tunisia" of the country parameter. The new preferences are weighted according to the assigned weight, always based on the hierarchical relationship. In our proposal, a father member in a hierarchy will have as weight the weighted average of the weights of its children.
Our method of interpreting preferences on members from preferences defined on child members takes into account the fact that the user is interested in the first place in the most important preferences (weight greater than φ). The calculation of the weight of the interpreted preference takes into account the number, and the weights of the preferences whose weight is greater than the personalization rate φ, in relation to the

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:01:51 UTC from IEEE Xplore. Restrictions apply.

total number of preferences expressed.
Moreover, the preference interpreted by our system emphasizes the hierarchical relationship between the members of two different levels: because the parent member is an aggregation of its child members. Thus, the attribution of an interest weight to a parent member implicitly reflects an interest in all its child members.
Given a dimension $D_i \in D\, and\, H^{Di}{}_j \in H^{Di}\, such\, as\, H^{Di}{}_j = \langle L^0{}_j, ..., L^m{}_j, all\rangle, \forall r \in [0, m] \bigwedge l \in dom(L^r{}_j) \bigwedge l^{'} \in dom(L^{r+1}{}_j) \bigwedge Weight(l) = u \bigwedge Weight_A^{Di}(L^{r+1}{}_j) > 0\, such\, as\, l^{'} = anc(l)$ with anc = ancestor :
If $(Weight_{parameter}(L^{r+1}{}_j = l^{'})) =\notin [0, 1]$, then we deduce that :
$Weight_{parameter}(L^{r+1}{}_j = l^{'}) = [S/P] * [P/Perf] + [(S^{'}/(Perf - P)] * [(Perf - P)/(N - P)]$ with :

- S is the sum of the preferences expressed on the child members greater than φ.
- $S^{'}$ is the sum of the preferences expressed on the child members less than φ.
- P is the number of preferences expressed on the child members greater than φ.
- Perf is the number of preferences expressed on child members.

5) Formalization of visualization constraints: A visualization constraint makes it possible to express a restriction on the number and the visualization structure of the data to be returned to the user [1]. The visualization of multidimensional data by cross tables can be defined by a regular format. This format expresses restrictions on the size of the data returned to the user. In our work, a visualization constraint Cv makes it possible to delimit not only the number of data returned to the user, by determining the number of preferred members of the dimensions, but also the number of levels of detail with which the data will be analyzed. A visualization constraint $C_v = \{(I^V, M^V), (I^h, M^h)\}$, with I^V and I^h representing the number of imbrications of the vertical and horizontal analysis axes, and M^V and M^h representing the number of members that we can have vertically and horizontally.

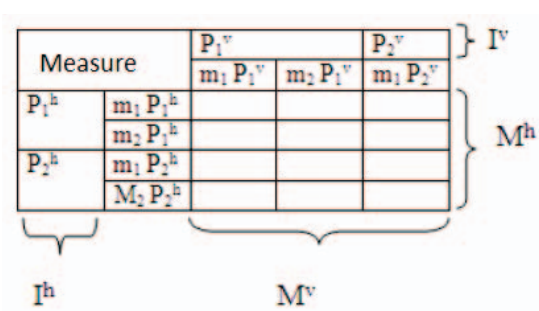


Fig. 2. Formalization of visualization constraints

V. Presentation of the PersonalizationDW tool

A. Purpose of PersonalizationDW

PersonalizationDW is a tool that performs a personalized OLAP query under constraint. It provides the following functionalities:

- Add users as new decision makers each of whom can make an MDX query.
- The creation of the elements of each user profile: the definition of the visualization constraint, the assignment of the weights associated with the multidimensional structures of each query.
- Return personalized query results based on user profile. The query is executed by a web service which connects to the ED located on a SQL Server DBMS. If the capacity of the devices used cannot display the entire result of the query, the system displays the data in an incremental manner according to the users' preferences stored in the profile.

B. Architecture of PersonalizationDW

The PersonalizationDW tool consists of four components:

- Query Expression: This module allows users to express their MDX queries.
- Personalization module: this module implements our hierarchical personalization method based on the user profile (the visualization constraints and preferences of multidimensional structures and data).
- Rewrite module: this module makes it possible to rewrite the initial query according to the user profile. Thus, we obtain so many personalized queries for the same initial query, as iterations.
- Data storage: our data warehouse documents are stored on the Sql Server DBMS.
- Metadata Repository: it stores the metadata describing the user's profile. This profile includes his preferences and the viewing constraint.

VI. Experimental study and results analysis

In order to validate our approach to personalizing OLAP queries, experiments were carried out to identify the performance of our personalization method by adopting our profile model. In this context, we compare the execution times of non-personalized queries with personalized queries. We present, in what follows, the results concerning the execution times of these queries. The implementation aims to evaluate the effectiveness of our approach to personalize OLAP queries integrating the user's profile in the query execution process. This evaluation consists of comparing the execution times of personalized queries and non-personalized queries.
The characteristics of our computer used for the evaluation tests are:

- Operating system: Windows 10.
- Processor: Intel(R) Core(TM) i3-10110U CPU @ 2.10GHz 2.59 GHz.
- Ram: 4 GB.

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:01:51 UTC from IEEE Xplore. Restrictions apply.

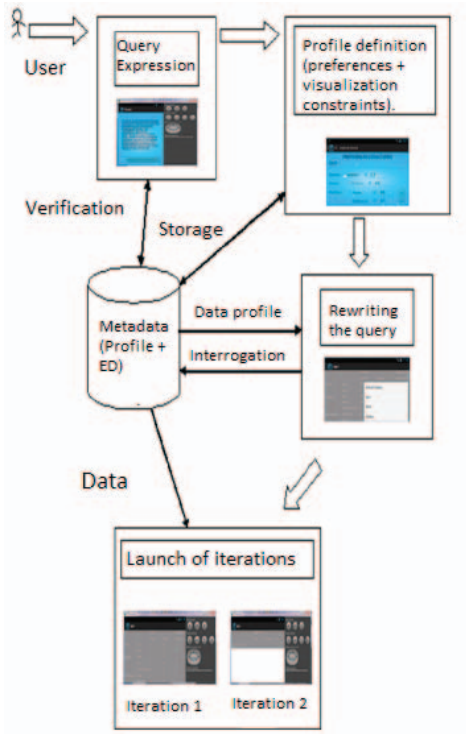


Fig. 3. PersonalizationDW Functional Architecture

In our experiments, we used the "Sales" cube and the user's preferences are expressed on 5 dimensions. The size of each table in terms of instances is described in the next table V. We took 5 queries that return 50, 170, 250, 190 and 210 cells.

Tables	Number of records
Customer	103
Product	104
Store	102
Sales Fact	300
Time	92

TABLE V
SIZE OF TABLES USED

We calculated their execution times on an emulator and not on a tablet. Then we personalized these queries according to user profiles and then we executed it. After our personalization step, we obtain an incremental result which is displayed in the form of 3 iterations (see table VI).

The following figure 4 shows the execution time savings of

Queries	Number of cells	Execution time (seconds)	
		Before personalization	After personalization
R1	50	4	3
R2	170	6	4
R3	250	8	6
R4	190	6	4
R5	210	7	5

TABLE VI
QUERY EXECUTION TIME

personalized queries compared to non-personalized queries. We notice that there is a significant gain in execution time

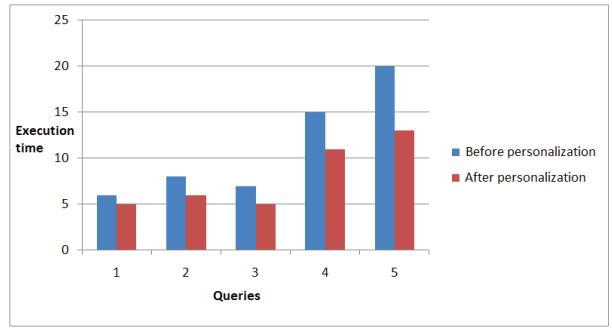


Fig. 4. runtime before and after personalization

for personalized queries with our personalization method than for non-personalized queries, this is especially seen when the initial query contains a large number of cells. Whatever the visualization constraint imposed by the user, the execution times of personalized queries always remain lower than the execution times of non-personalized queries thanks to our personalization method which represents the data in an incremental way as an iteration. Indeed, distributing the results over several iterations makes it possible to reduce the response time. This proves that our method of hierarchical personalization of data warehouses under visualization constraints acts positively on the execution time of user queries.

VII. CONCLUSION AND PROSPECTS

To achieve our objective, we have presented our method for personalizing decisional queries which takes into account the refinements requested by the user to perform his analyses. The personalized hierarchical visualization of the result is ensured even in limited display conditions. Indeed, our method makes it possible to adapt the result to the preferences of the profile, while taking into account the constraint of visualization. Moreover, our incremental method of personalizing the query result de nes the visualization constraint by giving the possibility to the user to have results according to all his preferences. Our method is based on a profile model allowing to personalize, following a quantitative approach, the concepts and the multidimensional data. This model is based on the hierarchical structure of the data in the warehouse to interpret new preferences not defined by the user. Several perspectives are to be considered in this work. In the short term, we propose to define the user profile in an implicit way. Thus, we can take advantage of OLAP query logs to induce user preferences by enriching these preferences with our interpretation method. At the level of the visualization constraint, we plan to enrich it with other constraints (memory constraint) detected automatically from the machine. In the long term, we consider the semantic aspect by enriching user queries with a domain ontology.

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:01:51 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Bellatreche, L., A. Giacometti, P. Marcel, H. Mouloudi, and D. Laurent, A Personalization Framework for OLAP Queries. DOLAP 2005, ACM 8th International Workshop on Data Warehousing and OLAP, Bremen, Germany, November 4-5, 2005, Proceedings: ACM, 2005.

[2] Favre, C., F. Bentayeb and O. Boussaid, Evolution et personnalisation des analyses dans les entrepôts de données - Une approche orientée utilisateur. Actes du XXVème Congrès INFORSID, Perros-Guirec, France, 2007.

[3] Jerbi, H., F. Ravat, O. Teste and G. Zurfluh, Management of context-aware preferences in multidimensional databases. Third IEEE International Conference on Digital Information Management (ICDIM), November 13-16, 2008, London, UK, Proceedings: IEEE, 2008.

[4] Ravat, F. et O. Teste, Personalization and OLAP Databases, New Trends in Data Warehousing and Data Analysis., vol. 3. Springer, 2009, pp.1–22.

[5] Agrawal, R., Rantzau, R., and Terzi, E, Context-sensitive ranking. Proceedings of the ACM SIGMOD International Conference on Management of Data, Chicago, Illinois, USA: ACM, 2006.

[6] Koutrika G., and Y. E. Ioannidis, Personalization of Queries in Database Systems. Proceedings of the 20th International Conference on Data Engineering, ICDE 2004, 30 March - 2 April 2004, Boston, MA, USA: IEEE Computer Society, 2004.

[7] Chomicki J., Querying with Intrinsic Preferences, Advances in Database Technology - EDBT 2002, 8th International Conference on Extending Database Technology, Prague, Czech Republic, March 25-27, Proceedings., vol. 2287. Springer, 2002, pp.34–51.

[8] Chomicki, J., "Preference formulas in relational queries," ACM Trans. Database Syst., vol. 20, pp. 137–157, 2004.

[9] Boutilier, C., Brafman, R I., Domshlak, C., Hoos, H H., Poole, D., "Preference-Based Constrained Optimization with CP-Nets," Comput. Intell., vol. 28, pp. 427–466, 2004.

[10] Boutilier, C., Brafman, R I., Domshlak, C., Hoos, H H., Poole, D., "CP-nets: A Tool for Representing and Reasoning with Conditional Ceteris Paribus Preference Statements," J. Artif. Intell. Res., vol. 21, pp. 135–191, 2004.

[11] Das, G., Hristidis, V., Kapoor, N., and Sudarshan, S., Ordering the attributes of query results. Proceedings of the ACM SIGMOD International Conference on Management of Data, Chicago, Illinois, USA: ACM, 2006.

[12] Favre, C., F. Bentayeb and O. Boussaid, Intégration des connaissances utilisateurs pour des analyses personnalisées dans les entrepôts de données évolutifs, Extraction et gestion des connaissances (EGC'2007), Actes des cinquièmes journées Extraction et Gestion des Connaissances, Namur, Belgique, 23-26 janvier 2007, 2 Volumes., vol. RNTI-E-9. Cépaduès-Éditions, 2007, pp.217–222.

[13] Bentayeb, F., Favre, C. and Boussaïd, O., "A user-driven data warehouse evolution approach for concurrent personalized analysis needs," Integrated Computer-Aided Engineering., vol. 15, pp. 21–36, 2008.

[14] Bentayeb, F., Boussaïd, O., Favre, C., Ravat, F. and Teste. O, Personnalisation dans les entrepôts de données : bilan et perspective, Actes des 5èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne, EDA 2009, Montpellier, France, vol. B-5. Cépaduès, 2009, pp.7–22.

[15] Khemyri, R. and Bentayeb, F, Utilisation des vues matérialisées pour la personnalisation des entrepôts de données, Actes des 5èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne, EDA 2009, Montpellier, France, vol. B-5. Cépaduès, 2009, pp.121–129.

[16] Garrigós, I., Pardillo, J., Mazón, J-N. and Trujillo, J., A Conceptual Modeling Approache for OLAP Personalization, Conceptual Modeling - ER 2009, 28th International Conference on Conceptual Modeling, Gramado, Brazil, November 9-12, 2009. Proceedings., vol. 5829. Springer, 2009, pp.401–414.

[17] Glorio, O., Mazón, J., Garrigós, I. and Trujillo, J., Using web-based personalization on spatial data warehouses. Proceedings of the 2010 EDBT/ICDT Workshops, Lausanne, Switzerland, March 22-26, 2010: ACM, 2010.

[18] Glorio, O., Mazón, J., Garrigós, I. and Trujillo, J., "A personalization process for spatial data warehouse development," Decision Support Systems., vol. 52, pp. 884–898, 2012.

[19] Mouloudi, H., Bellatreche, L., Giacometti, A. and Marcel, P., Personalization of MDX Queries. 22èmes Journées Bases de Données Avancées, BDA 2006, Lille, France, 17-20 octobre 2006, Actes (Informal Proceedings), 2006.

[20] Ravat, F., Teste, O. and Zurfluh, G., "Personnalisation de bases de données multidimensionnelles," Actes du XXVème Congrès INFORSID, Perros-Guirec, France, 22 au 25 mai 2007, 2007, pp. 121–136.

[21] Ravat, F. and Teste, O., Personalization and OLAP Databases, New Trends in Data Warehousing and Data Analysis., vol. 3. Springer, 2009, pp.1–22.

[22] Golfarelli, M. and Rizzi, S., Expressing OLAP Preferences, Scientific and Statistical Database Management, 21st International Conference, SSDBM 2009, New Orleans, LA, USA, June 2-4, 2009, Proceedings., vol. 5566. Springer, 2009, pp.83–91.

[23] Golfarelli, M., Rizzi, S. and Biondi, P., "myOLAP : An Approach to Express and Evaluate OLAP Preferences," IEEE Transactions on Knowledge and Data Engineering., vol. 23, pp. 1050–1064, 2011.

[24] Aligon, J., Golfarelli, M., Marcel, P., Rizzi, S., Turricchia, E., Mining Preferences from OLAP Query Logs for Proactive Personalization, Advances in Databases and Information Systems - 15th International Conference, ADBIS 2011, Vienna, Austria, September 20-23, 2011. Proceedings., vol. 6909. Springer, 2011, pp.84–97.

[25] Yang, Y., Abdelhédi, F., Darmont, J., Ravat, F. and Teste, O., "Automatic Machine Learning-Based OLAP Measure Detection for Tabular Data," Big Data Analytics and Knowledge Discovery - 24th International Conference, DaWaK 2022, Vienna, Austria, August 22-24, 2022, Proceedings, 2022, pp. 173–188.

[26] Bimonte, S., Gallinucci, E., Marcel, P. and Rizzi, S., "Logical design of multi-model data warehouses," Knowl. Inf. Syst., vol. 65, pp. 1067–1103, 2023.

[27] Francia, M., Golfarelli, M., Marcel, P., Rizzi, S. and Vassiliadis, P., "Assess Queries for Interactive Analysis of Data Cubes," Proceedings of the 24th International Conference on Extending Database Technology, EDBT 2021, Nicosia, Cyprus, March 23 - 26, 2021, 2021, pp. 121–132.

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:01:51 UTC from IEEE Xplore. Restrictions apply.

2. International Conference on Innovations in Intelligent Systems and Applications (INISTA 2023)

Article [C2] : Personalized OLAP queries under Hierarchical Visualization Constraint

Auteurs : **Hamdi Ghassen** and Omri Mohamed Nazih

Conférence : International Conference on Innovations in Intelligent Systems and Applications (**INISTA** 2023). 20-23 September 2023. Hammamet, Tunisia, pp. 1-6

Classement : **C**

Personalized OLAP queries under Hierarchical Visualization Constraint*

Publisher: IEEE

Cite This PDF

Ghassen Hamdi ; Mohamed Nazih Omri All Authors

17
Full
Text Views

Abstract

Document Sections

I. Introduction

II. Related Works

III. Data warehouse personalization

IV. Experimental study and results analysis

V. CONCLUSIONS

Authors

Figures

References

Keywords

Metrics

Abstract:
Decision makers do not evaluate all decision facts in the same way since their requirements differ greatly depending on their specialty, levels of knowledge, and other factors. Everyone has the right to have their own tastes and needs. The multidimensional model is intended for a group of decision-makers and consists of a set of data studied according to different dimensions. It contains all the indicators and analysis axes for a particular domain. They do not take into account the preferences of each decision-maker. A personalization phase of this diagram according to the preferences of these decision-makers would enrich the decision-making process. Another problem for consumers is the inability of mobile devices (smartphones, tablets, etc.) to display the results of OLAP queries. This paper presents a new approach in which preferences are expressed as visualization constraints that are applied sequentially to the schema to represent the result. This approach makes it possible to adapt the output to the profile choices while keeping in mind the constraint of visualization. Our strategy is based on the personalization of concepts and multidimensional data using a quantitative approach. This model interprets new preferences not expressed by the user based on the hierarchical structure of the data in the warehouse. We implemented our approach by considering three standard metrics, namely precision, recall, and F-measure. The results we obtained are very encouraging and confirm the feasibility of our proposed approach.

Published in: 2023 International Conference on Innovations in Intelligent Systems and Applications (INISTA)

Date of Conference: 20-23 September 2023

DOI: 10.1109/INISTA59065.2023.10310545

Date Added to IEEE *Xplore*: 22 November 2023

Publisher: IEEE

▸ **ISBN Information:**

Conference Location: Hammamet, Tunisia

▸ **ISSN Information:**

11/02/2024 22:53 INISTA 2023: notification for paper 1261 - hamdi.ghassan@gmail.com - Gmail

Gmail inista

99+
Mail
1
Chat
Meet

Nouveau message

Boîte de récep... 2080
Messages suivis
En attente
Important
Messages envoyés
Brouillons 120
Catégories
Plus

Libellés

Personnel
Unwanted
Plus

Dear Ghassen,

Congratulations! We are pleased to inform you that your submission:

- ID: 1261
- Title: Personalized OLAP queries under Hierarchical Visualization Constraint

has been accepted, as a REGULAR FULL PAPER, to INISTA 2023.

All papers were carefully reviewed by program committee members. The program committee worked diligently to reviev constructive review comments. They are attached at the end of this email, we hope you find them valuable, and ask yo and suggestions carefully when producing the camera-ready version of your paper.

NEXT STEPS:

1. Please immediately check that the author names and affiliations for all authors are correct in EasyChair, as these will forwarded to the IEEE Publisher (IEEE CPS). We remind you that alteration of authors' list and paper title are generally chairs.

2. The camera-ready version of the paper must be uploaded before August 15, 2023. You will receive soon by email de instructions for preparing the camera-ready version of your paper. The page limit for your camera-ready submission is 6 can be additionally purchased for the main text during registering your paper.

3. At least one author must register before August 15, 2023 with full fee paid as a presenter for the conference and com information will be available soon.

Once again, congratulations! We look forward to seeing you in Hammamet in September.

Best regards,
Conference chairs

ICORE Conference Portal

CORE rankings is now an international collaboration named ICORE

ICORE rankings page | Frequently asked questions

CORE Dora discussion document | Description of Conference Ranks | Guidelines for reporting rankings | Usage of CORE rankings | 2021 Rankings Process | Data used in Rankings and responses to common concerns

inista Search by: All Source: CORE2023

Search

Showing results 1 - 1 of 1

Export

CORE2023 Summary:
A* - 7.47% of 803 ranked venues
A - 14.45% of 803 ranked venues
B - 27.77% of 803 ranked venues
Australasian B - 0.75% of 803 ranked venues
C - 47.2% of 803 ranked venues
Australasian C - 2.37% of 803 ranked venues
Other - 164 total

Title	Acronym	Source	Rank	Note	DBLP	Primary FoR	Comments	Average Rating
International Symposium on Innovations in Intelligent Systems and Applications	INISTA	CORE2023	C	none	view	4602	0	N/A

2023 International Conference on Innovations in Intelligent Systems and Applications (INISTA) | 979-8-3503-3890-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/INISTA59065.2023.10310545

Personalized OLAP queries under Hierarchical Visualization Constraint*

Ghassen Hamdi
MARS Research Laboratory, University of Sousse.
Sousse, Tunisia
hamdighassan@gmail.com

Mohamed Nazih Omri
MARS Research Laboratory, University of Sousse.
Sousse, Tunisia
Mohamednazih.omri@eniso.u-sousse.tn

***Abstract*—Decision makers do not evaluate all decision facts in the same way since their requirements differ greatly depending on their specialty, levels of knowledge, and other factors. Everyone has the right to have their own tastes and needs. The multidimensional model is intended for a group of decision-makers and consists of a set of data studied according to different dimensions. It contains all the indicators and analysis axes for a particular domain. They do not take into account the preferences of each decision-maker. A personalization phase of this diagram according to the preferences of these decision-makers would enrich the decision-making process. Another problem for consumers is the inability of mobile devices (smartphones, tablets, etc.) to display the results of OLAP queries. This paper presents a new approach in which preferences are expressed as visualization constraints that are applied sequentially to the schema to represent the result. This approach makes it possible to adapt the output to the profile choices while keeping in mind the constraint of visualization. Our strategy is based on the personalization of concepts and multidimensional data using a quantitative approach. This model interprets new preferences not expressed by the user based on the hierarchical structure of the data in the warehouse. We implemented our approach by considering three standard metrics, namely precision, recall, and F-measure. The results we obtained are very encouraging and confirm the feasibility of our proposed approach.**

***Index Terms*—Data warehouse, OLAP queries, hierarchical visualization under constraints, personalization.**

I. Introduction

A. Context and issues

Decision-makers do not all visualize decision-making data in the same way because their needs vary greatly depending on their specialties, levels of knowledge, and other factors. Everyone is free to have their own preferences and requirements. For a group of decision-makers, the multidimensional model, which consists of a set of facts analyzed along various dimensions, includes all the indicators and axes of analysis relevant to a given domain. They do not take each decision-maker's preferences into account. The decision-making process will be enhanced by a phase of personalization of this diagram based on the preferences of these decision-makers. Another difficulty for users is the constrained capacity of the gadgets used (smartphones, tablets, etc.) to consult the outcomes of OLAP queries. In this situation, we suggest that the preferences be expressed as visualization constraints that will be applied to the diagram in order to incrementally represent the outcome. We first formalize the "User profile" concept for data warehouses in this article. Additionally, we describe how we personalize the user experience based on the incremental representation of the result. The display space restriction is taken into account in this visualization.

Many personalization works for data warehouses ignore the "hierarchy" concept, which characterizes the data structure in the data warehouse ([1], [2], [15], [22], [24]). These works prioritize the personalization of the data analysis results, ignoring the hierarchical structure in which the multidimensional data returned to the user are kept in the warehouse. The solutions suggested in the literature ([1], [16], [23]) are limited to returning the most important preferences while respecting the constraint, which is a barrier for the user who wants to view all the data. However, the limited capacity of the devices used to consult data from a data warehouse is a barrier for the user who wishes to view all the data. As a result, other preferences that are also significant are not included in the user's query's outcome.

B. Contributions

In this article, we suggest personalizing OLAP queries by visualizing data in accordance with the various levels of detail that users prefer. To have a more or less detailed overall vision (based on the levels of refinement requested), users of a data warehouse do in fact query results with varying degrees of refinement for specific axes of analysis. However, the users' access to this personalized hierarchical view of multidimensional data is limited and not gratuitous. It is constrained by the capacity of the device being used to display the results due to the data visualization constraint. Based on the preferences he has indicated in his profile, we present the user with a hierarchical view of the results. Additionally, we provide the user with an incremental visualization of the data returned by the personalized query, in a constrained context, to lessen the impact of the visualization constraint on the personalized results. Despite being in a context with a visualization constraint, our approach to personalizing decision queries is not limited to the visualization of dimensions with a single attribute ([1], [17]). In a constrained visualization environment, our method ensures users a hierarchical and incremental view of the data while taking into account their preferences for the multidimensional model's structures as well as data from the warehouse ([3], [4], [18]). In order

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:03:42 UTC from IEEE Xplore. Restrictions apply.

to accomplish this, we suggest a profile model that enables the expression of preferences for the multidimensional model and the warehouse data. Our profile model also enables the interpretation of new, undefined preferences by taking advantage of the hierarchical characteristic. In this paper, we extend OLAP queries written in the Microsoft MDX (Multi-Dimensional eXpressions) language.

This article's remaining sections are organized as follows. We outline the main strategies used to address the issue of personalization in the literature in section II. The motivation behind our personalization method will be presented in Section **??**. In Section III, we describe our suggested approach to personalizing data warehouses, which enables users to define their preferences by applying a quantitative approach to the multidimensional structures and data of the warehouse. The experimental study carried out in this work is provided, along with an analysis of the findings, in Section IV. We provide a summary of the work as well as some suggestions for upcoming projects in the final section V.

II. Related Works

In the area of databases, there are numerous personalization works available, but the personalization techniques used do not approach the issue of personalization in the same manner. This work can be categorized into two categories based on two different theoretical perspectives: the quantitative approach, which involves the user indirectly expressing their preferences by associating numerical scores generated by score functions, and the qualitative approach, where the preferences are directly specified by binary relations.

By using scoring functions and operating on the content, the personalization approach suggested by [5], [19] displays the preferences of the user profile. Constraints are not supported by this approach. In reality, the user profile merely contains preferences that determine the overall hierarchy of the tuples and act as filters for the query's responses. Following the calculation of the query result, the tuples are compared, and the user will only see those with notable interests. According to the profile's settings, this method shows the best tuples of a query's response.

A personalization technique based on the preferences of the user profile expressed on the material is suggested by the authors in [6], [20]. Selecting the user's preferences that are most pertinent to them is feasible using this method that is specified in a quantitative framework. For example, a score must be supplied to a selection predicate or a join predicate. These preferences are represented by scoring functions, which assign degrees of interest to individual query atomic items. These preferences then determine the overall ranking of the queries: if one query has a higher degree of interest than another, it is more intriguing overall.

In [7], [8], [27], the authors offer a personalization method that does not present constraints. On content, profile preferences are set. establishing binary connections between the query and response tuples. These preferences are expressed by a strict partial order or a binary relation. The preference operator WINNOW introduces them in the relational language, in [8]. As a result, a relational algebra query with preferences is one that uses the WINNOW operator at least once. Additionally, [8] introduces characteristics to enhance the calculation of chosen tuples from the calculated result of the query, i.e. following access to the data.

As a result, different scores can be attached to the same tuple. By changing the problem into one of selecting the most beneficial collection of attributes, or the Top-m attributes that have the greatest influence on the computation of the Top-n ordered tuples of the answer, with $m << M$ and $n << N$, the problem can be addressed. In their work, [2], [9], [10], [28] offer a personalization strategy that entails personalizing a data warehouse's schema by adding new hierarchical levels. The goal of this research is to develop methods for utilizing and sharing the knowledge of decision-makers that are stored in dimensions as combinations of new hierarchies in the form of "if-then" type rules, throughout the warehouse query process. The authors in [3] suggest a multidimensional database schema-based context-aware preference model. OLAP preferences are modeled using a qualitative method and are highly dependent on the user's analysis context. This is the reason a tree of OLAP analysis components is used to provide a conceptual model of the OLAP context. Regarding the same dimension, there are preferences between different dimensions.

A personalization technique based on the creation of preferences is put forth by [11], [12], [26] in order to personalize the analyses. The multidimensional notions (fact, dimension, hierarchy, measurements, parameter, or attribute) form the basis of the suggested conceptual model. A set of user preferences given as a priority order on the attributes of the multidimensional model is used to alter the contents of the multidimensional tables.

According to their work, [13], [14], [21], [25] suggest personalizing the wording of MDX queries to the user's choices. With this method, preferences can be expressed for the OLAP query. A proposed OLAP algebra called MYOLAP allows users to specify their preferred attributes, measurements, and level of aggregation.

III. Data warehouse personalization

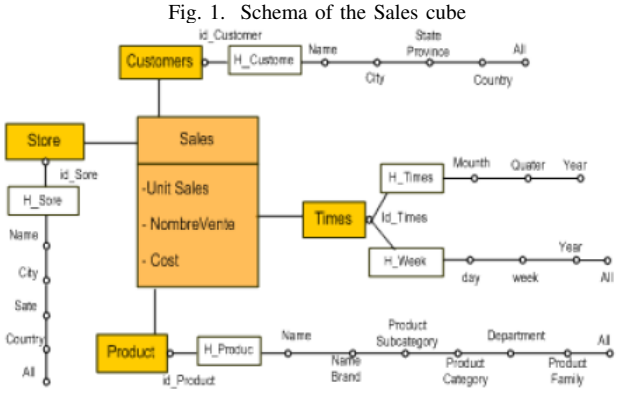
We will use the query below as an example, where the user is querying the Sales cube, whose schema is displayed in Figure 1.

The following query represents the user's desire to see the measurements (NumberSale) broken down into the four levels of the Product dimension (Family, Department, Category, Subcategory) and the three levels of the Customers dimension (Country, State, Province, City).

"$SELECT\ \{[customer].[hcustomer].[country],$
$[customer].[hcustomer].[stateprovince],$
$[customer].[hcustomer].[city]\}ONCOLUMNS,$
$\{[product].[HProduct].[family].members,$
$[product].[HProduct].[category].members,$
$[product].[HProduct].[subcategory].members\}ONROWS$
$FROM[SalesCube]where([Measures].[unitsales])$"

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:03:42 UTC from IEEE Xplore. Restrictions apply.

Fig. 1. Schema of the Sales cube



An environment known as a personalized data warehouse is one where user profiles and data's multidimensional properties are combined. A multidimensional conceptual model, profiles, and preference expression functions describe a personalized data warehousing Ep. :
$Ep = (Sch(C), P^C, F^C)$ with: Sch(C) : represents the multidimensional schema of the data warehouse: the constellation schema.
We use the constellation model which groups together several subjects of analysis (facts) studied according to different axes of analysis (dimensions) which may be shared. These dimensions are composed of several hierarchies.
$P^C = \{P_1, ..., P_u\}$: represents the list of warehouse user profiles.
$F^C = \{f_1, ..., f_z\}$: is the set of functions for personalizing the data warehouse and setting user preferences.

In this work, we suggest using a quantitative technique to define user preferences by personalizing the multidimensional structures and the data of the warehouse. In our scoring functions, users can apply importance weights between 0 and 1, which indicate how interested they are in a certain structure or set of data.

A. The representation of the profile

The profile in the context of our work comprises viewing preferences and limits. Preferences are represented on both data (instances) and multidimensional structures (measures, dimensions, hierarchies, etc.). The number of result data given is limited by the visualization limitation, which is connected to the user's tool for querying the data cube.

A profile P can be defined by: $P = (Pr_S, Pr_D, C_v)$ with:

- Pr_S is the preferences expressed on multidimensional structures.
- Pr_D is the set of preferences expressed on the instances.
- C_v represents the visualization constraint characterizing the context of visualization of the results for the user.

1) The formalization of preferences: Decision makers can define their preferences by utilizing functions of F, which are generally based on scoring functions and assigning a value between 0 and 1. The multidimensional structures and the data both express the preferences that make up a profile.

2) Structure Preferences: The user can define his preferences on the multidimensional structures he wishes to visualize in the results of the queries, by assigning degrees of importance. Let $Weight_z \in F$, a function called: $Weight_z : Z \rightarrow [0, 1]$, which assigns a weight between 0 and 1 to a multidimensional structure $z \in Z, Z \subseteq Sch(C)$.
Preferences are expressed on dimensions, hierarchies, measures, aggregation functions, etc. by assigning weights between 0 and 1 using the function $Weight_z$.
Consider the example of setting preferences on dimensions, using the function $Weight_D$.
$Weight_D : D \rightarrow [0, 1] \ \forall i \in [1, n] \, and \, D_i \in D, Weight_D(D_i) \in [0, 1]$

$D = \{Customers, Product, Times, Store\}$.
$Weight_D(Customers) = 0.65; Weight_D(Product) = 0.9; Weight_D(Times) = 0.85; Weight_D(Store) = 0.75$

The function $Weight_D(D_i)$ provides a general weight on a dimension without specifying the weight of these sub-elements (hierarchies, levels). These factors must be taken into consideration for the ultimate weight that can be assigned to a dimension; for this reason, we create the concept of intra-dimension weight.

Extra dimension weight is a second weight that is described to define an order relation between elements of several dimensions. As a result, we distinguish between an extra-dimensional weight (noted by PE) and an intra-dimensional weight (noted by PI) when it comes to the weights given to the OLAP preferences specified on the analysis axes.

a) Intra dimension weight: An intra-dimension weight is assigned by the user to the hierarchies and analysis parameters of a dimension. This score only compares between hierarchies or parameters within the same dimension.
The Weight function defined previously is used to declare these weights.

- The intra-dimension weight of a hierarchy: The user can give importance weights to the hierarchies of a dimension $D_i \in D$ using the function: PI^{Di} :

$$H^{Di} \rightarrow [0, 1] \ \forall H_j^{Di} \in H^{Di}, PI_{H^{Di}}(H^{Di}j) \in [0, 1].$$

  Consider the Times dimension:
  $PI_{H^{Times}}(H_Times) = 0.9$; $PI_{H^{Times}}(H_week) = 0.3$
- The intra-dimension weight of a parameter: The user can define his preferences on the parameters with which he prefers to see the results. This using the function:

$$PI_{A^{Di}} : A^{Di} \rightarrow [0, 1] \ \forall a \in A^{Di}, PI_{A^{Di}}(a) \in [0, 1].$$

We take the Times dimension as an example:
$A^{Times} = \{PI_{A^{Times}}(idTimes) = 0.1; PI_{A^{Times}}(Day) = 0.2; PI_{A^{Times}}(Mounth) = 0.75; PI_{A^{Times}}(week) = 0.85; PI_{A^{Times}}(Year) = 0.95; PI_{A^{Times}}(all) = 0.5\}$

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:03:42 UTC from IEEE Xplore. Restrictions apply.

b) Extra dimension weight: An extra dimension weight, denoted by PE, is calculated based on the PI and the weight of the dimension. It is used to indicate the importance of hierarchies and parameters of different dimensions to compare them with each other.

- The extra dimension weight of a hierarchy: The PE of a hierarchy is calculated according to its PI and the weight of the dimension to which it belongs. $\forall i \in [1,n]\, and\, D_i \in D, \forall H_j^{Di} \in H^{Di} : PE(H_j^{Di}) = PI(H_j^{Di}) * Weight_D(Di)$.

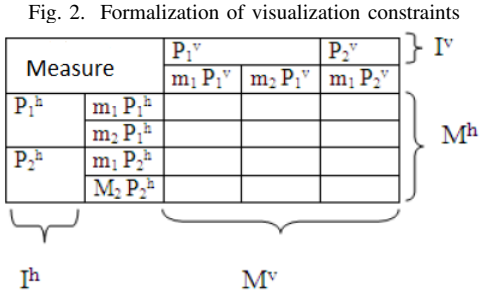
We take the hierarchies from the Times dimension.
$PE(H_Times) = PI_{H^{Times}}(H_Times) * Weight_D(Times) = 0.9 * 0.85 = 0.765$;
$PE(H_week) = 0.255$

- The extra dimension weight of a parameter: The PE of an analysis parameter depends not only on its PI and the weight of the dimension but also on the maximum PI of the hierarchies to which it belongs; since a parameter can belong to one or more hierarchies in a dimension. $\forall a \in A^{Di}, let\, H_a = \{H_1, ..., H_b\}\, such\, as\, \forall j \in [1,b], a \in H_j \wedge H_j \in H^{Di} : PE(a) = PI_{A^{Di}}(a) * max(\{PI_{H^{Di}}(H_j)\ \ j \in [1,b]\}) * Weight_D(D_i) = PI_{A^{Di}}(a) * max(\{PE(H_j)\ j \in [1,b]\})$.

$PE(Year) = PI_{A^{Time}}(Year) * max(\{PE(H_week), PE(H_Times)\}) = 0.95 * 0.765 = 0,727$.

3) Formalization of visualization constraints: A visualization constraint allows you to specify limitations on the quantity and type of data that will be returned to the user ([1]). Cross-table visualization of multidimensional data can be described by a standard format. The size of the data sent to the user is limited by this format. In our work, a visualization constraint Cv allows us to limit both the amount of data returned to the user and the number of levels of detail at which the data will be evaluated by identifying the preferred members of the dimensions. A visualization constraint $C_v = \{(I^V, M^V), (I^h, M^h)\}$, with I^V and I^h representing the number of the imbrication of the vertical and horizontal analysis axes, and M^V and M^h representing the number of members that we can have vertically and horizontally (see figure 2).

Fig. 2. Formalization of visualization constraints



IV. Experimental study and results analysis

A. Description of test corpora

The client submits an MDX query to the personalization system, which the personalization system then must personalize. The best subset of dimensions, parameters, and members should be retained for a given profile. This is the basic idea of personalisation.
Experiments were conducted to determine the performance of our personalization method by implementing our profile model in order to validate our approach to personalizing OLAP queries.
The characteristics of our computer used for the evaluation tests are:

- Operating system: Windows 10.
- Processor: Intel(R) Core(TM) i3-10110U CPU @ 2.10GHz 2.59 GHz.
- Ram: 4 GB.

In our experiments, we used the "Sales" cube and the user's preferences are expressed on 5 dimensions. The size of each table in terms of instances is described in the next table. The fundamental performance indicators utilized to assess

TABLE I
Size of tables used

Tables	Number of records
Customer	103
Product	104
Store	102
Sales Fact	300
Time	92

our approach are of interest to us. In this instance, our approach divides the data warehouse's instances into two categories: relevant cells and irrelevant cells. True positives are cells that the system correctly classifies as belonging to the category "relevant," while false positives are cells that the system incorrectly classifies. "True negatives" refer to cells that the system correctly categorized as belonging to the class "irrelevant"; "false negatives" refer to cells that were incorrectly categorized.

An expert in the field (researcher, linguist, engineer, etc.) evaluates the significance of the retrieved cells and determines whether they make sense by applying a weight (value between 0 and 1) to their dimensions, parameters, and members.
The main objective of this method is to increase the rate of genuine positives and decrease the rate of false positives, while simultaneously maximizing the number of real positives. We can then create a Decision Matrix (II) that compiles all the criteria required to assess the effectiveness of a binary classification system.

Thus, We compute the four following performance indices:

- True Positives (TP): Number of correctly classified positive cells.
- False Negatives (FN): Number of incorrectly classified positive cells.

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:03:42 UTC from IEEE Xplore. Restrictions apply.

- True Negatives (TN): Number of correctly classified negative cells.
- False Positives (FP): Number of incorrectly classified negative cells.

TABLE II
DECISION MATRIX

	Relevant cells	Irrelevant cells
Cells evaluated as relevant by the system	True Positives (TP)	False Positives (FP)
Cells evaluated as irrelevant by the system	False Negatives (FN)	True Negatives (TN)

B. Evaluation measures

In the following, we present the precision, recall, F-measure, Accuracy and G-mean measures ([29], [30]), which we will use to evaluate the performance of our approach:

- Precision (P): The ratio of the number of relevant cells by the total extracted cells.

$$P = \frac{TP}{TP + FP} \tag{1}$$

- Recall (R): The ratio of the number of extracted cells by the total number of relevant cells.

$$R = \frac{TP}{TP + FN} \tag{2}$$

- F-measure (F) is the harmonic average of the precision P and the recall R:

$$F = \frac{2 * (P * R)}{(P + R)} \tag{3}$$

C. Evaluations and results

We took 5 queries that return 50, 170, 250, 190, and 210 cells. Table III and figures 3, 4 show the precision, recall, and F-measure measures for our approach compared to the approach presented in [7], [8], [27]. Our approach proposes a personalization method of OLAP queries under hierarchical visualization constraints while the other system [7], [8], [27] proposes a personalization method without taking into account the visualization constraint, after launching five queries on our data warehouse. All irrelevant cells recovered constitute what is called ”the noise”. Precision is opposed by cell noise. If it is high, this indicates that few unnecessary cells are offered by the system and that the system can be considered ”precise”.

TABLE III
PRECISION, RECALL, AND F-MEASURE MEASURES FOR 5 QUERIES EXPRESSED IN %.

Queries	Personalization without visualization constraint Chomicki et al., (2002, 2003 and 2018)			Personalization under visualization constraint		
	P	R	F	P	R	F
Q1	52	47	49.37	78	65	70.90
Q2	61.45	68.2	64.65	83.52	82.08	82.79
Q3	62.5	64.4	63.43	85.6	86.29	85.94
Q4	56.4	57.3	56.84	83.15	80.20	81.64
Q5	53.7	58.5	56	81.9	78.89	80.36

Fig. 3. Precision, recall and F-measure measures for 5 queries expressed in % (Personalization without visualization constraint Chomicki et al., ([7], [8], [27]))

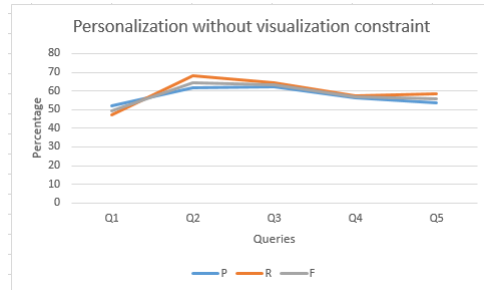
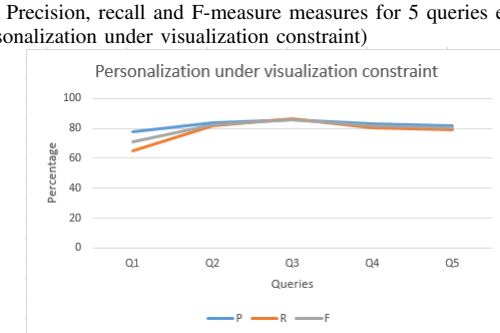


Fig. 4. Precision, recall and F-measure measures for 5 queries expressed in % (Personalization under visualization constraint)



D. Discussion

By analyzing the results obtained in the series of experiments, we notice the performance of our approach. The precision, recall, and f measurement shown in table III and figures 3, 4 show that our approach is very efficient in terms of productivity, ie in terms of the number of relevant cells rendered.

V. CONCLUSIONS

To accomplish our goal, we offered our method for personalizing decisional queries, which takes into account the modifications requested by the user when performing his analysis. Even in constrained display settings, customized hierarchical visualization of the result is ensured. Indeed, our strategy allows us to tailor the output to the profile's choices while keeping the visualization limitation in mind. Furthermore, by allowing the user to obtain results based on all of his preferences, our incremental way of personalizing the query result overcomes the visualization constraint. Our strategy is based on a profile model that allows us to personalize concepts and multidimensional data using a quantitative approach. This model interprets new preferences not expressed by the user based on the hierarchical structure of the data in the warehouse. This work will take into account a variety of viewpoints. In the short run, we propose that the user profile be defined implicitly. As a result, we may leverage OLAP query logs to instill user preferences by enhancing these preferences with our interpretation mechanism. We intend to supplement the visualization constraint with other constraints (memory

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:03:42 UTC from IEEE Xplore. Restrictions apply.

constraints) discovered automatically by the machine. We will examine the semantic aspect in the long run by enhancing user queries with a domain ontology.

REFERENCES

[1] Bellatreche, L., A. Giacometti, P. Marcel, H. Mouloudi, and D. Laurent, A Personalization Framework for OLAP Queries. DOLAP 2005, ACM 8th International Workshop on Data Warehousing and OLAP, Bremen, Germany, November 4-5, 2005, Proceedings: ACM, 2005.

[2] Favre, C., F. Bentayeb and O. Boussaid, Evolution et personnalisation des analyses dans les entrepôts de données - Une approche orientée utilisateur. Actes du XXVème Congrès INFORSID, Perros-Guirec, France, 2007.

[3] Jerbi, H., F. Ravat, O. Teste, and G. Zurfluh, Management of context-aware preferences in multidimensional databases. Third IEEE International Conference on Digital Information Management (ICDIM), November 13-16, 2008, London, UK, Proceedings: IEEE, 2008.

[4] Ravat, F. et O. Teste, Personalization and OLAP Databases, New Trends in Data Warehousing and Data Analysis, vol. 3. Springer, 2009, pp.1–22.

[5] Agrawal, R., Rantzau, R., and Terzi, E, Context-sensitive ranking. Proceedings of the ACM SIGMOD International Conference on Management of Data, Chicago, Illinois, USA: ACM, 2006.

[6] Koutrika G., and Y. E. Ioannidis, Personalization of Queries in Database Systems. Proceedings of the 20th International Conference on Data Engineering, ICDE 2004, 30 March - 2 April 2004, Boston, MA, USA: IEEE Computer Society, 2004.

[7] Chomicki J., Querying with Intrinsic Preferences, Advances in Database Technology - EDBT 2002, 8th International Conference on Extending Database Technology, Prague, Czech Republic, March 25-27, Proceedings., vol. 2287. Springer, 2002, pp.34–51.

[8] Chomicki, J., "Preference formulas in relational queries," ACM Trans. Database Syst., vol. 20, pp. 137–157, 2004.

[9] Favre, C., F. Bentayeb and O. Boussaid, Intégration des connaissances utilisateurs pour des analyses personnalisées dans les entrepôts de données évolutifs, Extraction et gestion des connaissances (EGC'2007), Actes des cinquièmes journées Extraction et Gestion des Connaissances, Namur, Belgique, 23-26 janvier 2007, 2 Volumes., vol. RNTI-E-9. Cépaduès-Éditions, 2007, pp.217–222.

[10] Bentayeb, F., Favre, C. and Boussaïd, O., "A user-driven data warehouse evolution approach for concurrent personalized analysis needs," Integrated Computer-Aided Engineering., vol. 15, pp. 21–36, 2008.

[11] Ravat, F., Teste, O. and Zurfluh, G., "Personnalisation de bases de données multidimensionnelles," Actes du XXVème Congrès INFORSID, Perros-Guirec, France, 22 au 25 mai 2007, 2007, pp. 121–136.

[12] Ravat, F. and Teste, O., Personalization and OLAP Databases, New Trends in Data Warehousing and Data Analysis., vol. 3. Springer, 2009, pp.1–22.

[13] Golfarelli, M. and Rizzi, S., Expressing OLAP Preferences, Scientific and Statistical Database Management, 21st International Conference, SSDBM 2009, New Orleans, LA, USA, June 2-4, 2009, Proceedings., vol. 5566. Springer, 2009, pp.83–91.

[14] Golfarelli, M., Rizzi, S. and Biondi, P., "myOLAP : An Approach to Express and Evaluate OLAP Preferences," IEEE Transactions on Knowledge and Data Engineering., vol. 23, pp. 1050–1064, 2011.

[15] Hamdi, G., Omri, M. N., Papini, O., Benferhat, S. and Bouraoui, Z., Querying DL-lite Knowledge Bases from Hidden Datasets. International Symposium on Artificial Intelligence and Mathematics, ISAIM 2018, Fort Lauderdale, Florida, USA, January 3-5, 2018, 2018.

[16] Hamdi, G. and Omri, M. N., New Model for Handling Inconsistency Issues in DL-Lite Knowledge Bases, New Trends in Intelligent Software Methodologies, Tools and Techniques - Proceedings of the 20th International Conference on New Trends in Intelligent Software Methodologies, Tools and Techniques, SoMeT 202, Cancun, Mexico, 21-23 September, 2021., vol. 337. IOS Press, 2021, pp.183–196.

[17] Hamdi, G. and Omri, M. N., Towards New Model for Handling Inconsistency Issues in DL-Lite Knowledge Bases, Database and Expert Systems Applications - 32nd International Conference, DEXA 2021, Virtual Event, September 27-30, 2021, Proceedings, Part II., vol. 12924. IOS Press, 2021, pp.93–99.

[18] Hamdi, G., Omri, M. N., Benferhat, S., Bouraoui, Z., and Papini, O., "Query answering DL-lite knowledge bases from hidden datasets," Ann. Math. Artif. Intell., vol. 89, pp. 271–299, 2021.

[19] Hamdi, G., Telli, A. and Omri, M. N., "Querying of several DL-Lite knowledge bases from various information sources-based polynomial response unification approach," J. King Saud Univ. Comput. Inf. Sci., vol. 34, pp. 1161–1171, 2022.

[20] Telli, A., Hamdi, G. and Omri, M. N., "Lexicographic Repair Under Querying Prioritised DL-Lite Knowledge Bases," Scientific Journal of King Faisal University / Basic and Applied Sciences., vol. 22, 2021.

[21] Francia, M., Gallinucci, E., Golfarelli, M. and Rizzi, S., Insight-Based Vocalization of OLAP Sessions, Advances in Databases and Information Systems - 26th European Conference, ADBIS 2022, Turin, Italy, September 5-8, 2022, Proceedings., vol. 13389. Springer, 2022, pp.193–206.

[22] Bouhatous, A., Bellatreche, L., Abdelwahed, E. H., Ordonez, C., "The Impact of Multicore CPUs on Eco-Friendly Query Processors in Big Data Warehouses," IEEE International Conference on Big Data, Big Data 2022, Osaka, Japan, December 17-20, 2022, 2022, pp. 4463–4472.

[23] Benkrid, S., Bellatreche, L., Mestoui. Y. and Ordonez, C., "PROADAPT: Proactive framework for adaptive partitioning for big data warehouses," Data Knowl. Eng., vol. 142, pp. 102102, 2022.

[24] Azri, A., Favre, C., Harbi, N., Darmont, J. and Noûs, C., "MONITOR: A Multimodal Fusion Framework to Assess Message Veracity in Social Networks," Advances in Databases and Information Systems - 25th European Conference, ADBIS 2021, Tartu, Estonia, August 24-26, 2021, Proceedings, 2021, pp. 73–87.

[25] Francia, M., Golfarelli, M., Marcel, P., Rizzi, S. and Vassiliadis, P., "Assess Queries for Interactive Analysis of Data Cubes," Proceedings of the 24th International Conference on Extending Database Technology, EDBT 2021, Nicosia, Cyprus, March 23 - 26, 2021, 2021, pp. 121–132.

[26] Yang, Y., Abdelhédi, F., Darmont, J., Ravat, F. and Teste, O., "Automatic Machine Learning-Based OLAP Measure Detection for Tabular Data," Big Data Analytics and Knowledge Discovery - 24th International Conference, DaWaK 2022, Vienna, Austria, August 22-24, 2022, Proceedings, 2022, pp. 173–188.

[27] Chomicki, J., "Preference Queries," Encyclopedia of Database Systems, Second Edition, 2018.

[28] Attaf, S., Benblidia, N. and Boussaid, O., "Warehousing and analysing textual data," Int. J. Math. Model. Numer. Optimisation., vol. 8, pp. 216–244, 2018.

[29] Manning, C. D. and Schütze, H., "Foundations of statistical natural language processing," MIT Press, 2001.

[30] Raghavan, V. V., Jung, G. S., and Bollmann, P., "A Critical Investigation of Recall and Precision as Measures of Retrieval System Performance," ACM Trans. Inf. Syst., vol. 7, pp. 205–229, 1989.

Authorized licensed use limited to: UNIVERSITE DE SOUSSE. Downloaded on February 12,2024 at 13:03:42 UTC from IEEE Xplore. Restrictions apply.

3. **International Conference on Database and Expert Systems Applications (DEXA 2021)**

Article [C3] : Towards New Model for Handling Inconsistency Issues in DL-Lite Knowledge Bases

Auteurs : **Hamdi Ghassen** and Omri Mohamed Nazih

Conférence : The 32nd International Conference on Database and Expert Systems Applications (**DEXA** 2021), 27-30 September 2021. Virtual Event, pp. 93-99

Classement : **B**

SPRINGER LINK

Account

Find a journal | Publish with us | Track your research | Search | Cart

International Conference on Database and Expert Systems Applications

↳ DEXA 2021: **Database and Expert Systems Applications** pp 93–99 | Cite as

Home > Database and Expert Systems Applications > Conference paper

Towards New Model for Handling Inconsistency Issues in DL-Lite Knowledge Bases

Ghassen Hamdi & Mohamed Nazih Omri

Conference paper | First Online: 01 September 2021

695 Accesses

Part of the Lecture Notes in Computer Science book series (LNISA,volume 12924)

Download book PDF

Download book EPUB

Sections | References

Abstract

Introduction

ICORE Conference Portal

CORE rankings is now an international collaboration named ICORE

ICORE rankings page | Frequently asked questions

CORE Dora discussion document | Description of Conference Ranks | Guidelines for reporting rankings | Usage of CORE rankings | 2021 Rankings Process | Data used in Rankings and responses to common concerns

dexa Search by: All Source: CORE2021

Search

Showing results 1 - 1 of 1

Export

CORE2021 Summary:
A* - 7.45% of 792 ranked venues
A - 16.41% of 792 ranked venues
B - 36.74% of 792 ranked venues
Australasian B - 1.39% of 792 ranked venues
C - 36.24% of 792 ranked venues
Australasian C - 1.77% of 792 ranked venues
Other - 164 total

Title	Acronym	Source	Rank	Note	DBLP	Primary FoR	Comments	Average Rating
International Conference on Database and Expert Systems Applications	DEXA	CORE2021	B	none	view	4605	0	N/A

Towards New Model for Handling Inconsistency Issues in DL-Lite Knowledge Bases

Ghassen Hamdi[1] and Mohamed Nazih Omri[1]

MARS Research Laboratory, University of Sousse, Tunisia
hamdighassan@gmail.com
mohamednazih.omri@eniso.u-sousse.tn

Abstract. The lightweight description logic (DL-lite) represents one of the most important logic specially dedicated to applications that handle large volumes of data. Managing inconsistency issues, in order to effectively query inconsistent DL-Lite knowledge bases, is a topical issue. Since assertions (ABoxes) come from a variety of sources with varying degrees of reliability, there is confusion in hierarchical knowledge bases. As a consequence, the inclusion of new axioms is a main factor that causes inconsistency in this type of knowledge base. Often, it is too expensive to manually verify and validate all assertions. In this article, we study the problem of inconsistencies in the DL-Lite family and we propose a new algorithm to resolve the inconsistencies in prioritized knowledge bases. We carried out an experimental study to analyze and compare the results obtained by our proposed algorithm, in the framework of this work, and the main algorithms studied in the literature. The results obtained show that our algorithm is more productive than the others, compared to standard performance measures, namely precision, recall and F-measure.

Keywords: DL-Lite · Ontology · Inconsistency · Prioritized Knowledge Bases.

1 Introduction

DL-Lite [1] is a family of tractable DLs designed for applications that deal with large quantities of data and where the most relevant reasoning function is to respond to queries. DL-Lite ensures a low level of computational complexity and it is considered as especially well suited to Ontology-Based Data Access (OBDA) [2]. To reflect generic knowledge, Description logic (DL) use only concepts and role inclusions in logical formulas (called axioms) . A DL knowledge base (KB) is divided into two parts: a terminological base (called TBox), considered "generic" or "global", true in all models and for all individuals, and an assertional base (called ABox), this information is "specific" or "local", true for certain particular individuals.

How to deal with inconsistency is a crucially important question that emerges in OBDA ([3,4]. With respect to some assertions that contradict the terminology, inconsistency is defined in such a setting. Usually, a TBox is typically checked and confirmed, while the assertions may be supplied by different and inaccurate sources in vast amounts and can contradict the TBox. Furthermore, manually checking and validating all of the assertions is often prohibitively expensive. This is why, in OBDA, reasoning in

the face of inconsistency is important. Several works (*e.g.*[5]), inspired fundamentally by database approaches or propositional logic approaches treat with inconsistency in DLs , by adapting many inconsistency tolerant inference techniques, named semantics. These are focused on the assertional (or ABox) repair principle, which is closely related to the maximally consistent subset concept used in propositional logic . A repair of an ABox is merely a maximum assertional subbase that is consistent with a given TBox.

In this paper, the key current inconsistency-tolerant reasoning methods for prioritized KBs are first discussed. Interestingly enough, our approach is appropriate for the DL-Lite environment in the sense that, by producing a single preferred assertional repair, they allow effective handling of inconsistency.

The remainder of this article is arranged as follow. In section 2 we present the main approaches that tried to solve the problem of repairing Inconsistencies in Dl-Lites Knowledge Bases. Section 3 will be dedicated to the presentation of an overview L-Lite ontology and management of inconsistencies. Section 4 details our proposed approach and presents the used algorithm. It also provides the experimental study conducted in this work and analyse the obtained results. In the final section 5, we present a summary of the work as well as some ideas for future projects.

2 Related Works

Several works (*e.g.* [5]) attempted to deal with DL-Lite inconsistency by adapting several inference methods that accept inconsistency, based on database approaches (*e.g.* [6]). Priorities play an important role in the management of inconsistency, and they have been extensively discussed in the literature in the sense of propositional logic (*e.g.*[7]). In querying inconsistent databases or DL KBs, numerous works examined the notion of priority (*e.g.*[8]). Regretfully, only a few works are available in the OBDA format, such as the one given in [10,11] for dealing with reasoning under the prioritized DL-Lite ABox. In a recent line of research, the inconsistency in lightweight ontology is being investigated. In particular, the writers [5], look at the issue of KB inconsistency by computing a collection of consistent subsets of assertions known as repairs, which recover ontology consistency, and then using them to address queries. In addition, in [13,14], polynomials algorithms are proposed by the authors to pick a single preferred repair from a prioritized inconsistent DL-Lite KB, allowing for efficient query answering once the repair has been decided. Specifically, the writers in [13] propose a new approach focused on a single preferred repair solution. However, sequential inference strategies based on the selection of a single coherent assertional basis are proposed by the authors in [14]. In [15,16,17], the authors propose a new algorithm for answering queries without requiring access to Web databases.

3 DL-Lite ontology and management of inconsistencies: an overview

Inconsistency Treatment for Prioritized DL-Lite Assertional Bases: A DL-Lite knowledge base with priority is the set of assertions denoted by: $\mathcal{A} = \{S_1, ..., S_n\}$. The sets S_i

are called layers and each layer S_i contains the set of assertions with the same priority level i and are considered to be the most reliable than those present in a layer S_j when j> i. As a consequence, S_1 includes the most relevant assertions, while S_n contains the less important assertions. In the following, we use the DL-Lite notation $\mathcal{K}$=$\langle\mathcal{T},\mathcal{A}\rangle$ to link to a prioritized DL-Lite knowledge base of the kind: $\mathcal{A} = \{S_1, ..., S_n\}$.

Linear-Based Repair: Let $\mathcal{K}$=$\langle\mathcal{T},\mathcal{A}\rangle$ be an inconsistent and priority DL-Lite knowledge base. The linear based repair of $\mathcal{A}$ [14] is indicated by: $\ell(\mathcal{A}) = \{S'_1, ..., S'_n\}$ such that

- i) if i=1 then

$$S_1^{'} = \begin{cases} Sl & \text{if } \langle\mathcal{T}, S_l\rangle \text{ is consistent} \\ \emptyset & Otherwise \end{cases} \tag{1}$$

- ii) for i = 2, ..., n

$$S_i^{'} = \begin{cases} S_i & \text{if } \left\langle\mathcal{T}, S'_1 \cup \ldots \cup S'_{(i-1)} \cup S_i\right\rangle \text{ is consistent} \\ \emptyset & Otherwise \end{cases} \tag{2}$$

Clearly,$\ell(\mathcal{A})$ is acquired by dismissing an S_i layer when its facts contradict with the previous layers. Indeed, the subbase $\ell(\mathcal{A})$ is unique and it is consistent with $\mathcal{T}$.

Non-Defeated Repair: Another way to achieve a preferred repair is to interractively recover, layer by layer, all of the free elements free ($\mathcal{A}$).

We consider $\mathcal{K}$=$\langle\mathcal{T},\mathcal{A}\rangle$ a DL-Lite knowledge base with priority. We denote by *free*($\mathcal{A}$) the set of assertions relate to A that aren't to blame for any inconsistencies in $\langle\mathcal{T},\mathcal{A}\rangle$.

The Non-Defeated Repair [14], denoted by $nd(\mathcal{A})$ is a sequence $nd(\mathcal{A}) = S'_1 \cup ... \cup S'_n$, such that:

$$\forall i = 1 \ldots n, S_i^{'} = free(S_1 \cup \ldots \cup S_i)$$

Namely, $nd(\mathcal{A}) = free(S_1) \cup free(S_1 \cup S_2) \cup \ldots \cup free(S_1 \cup \ldots \cup S_{S_i})$.

4 Most-Possible Repair proposed approach

4.1 Most-Possible Repair Algorithm

Our algorithm is presented in this way: First, we initialize M-PR($\mathcal{A}$) on an empty set. Then we sort the layers according to their cardinalities in a decreasing way (according to the layers that contain the largest number of assertions). If two sources have the same cardinality, we apply FIFO (First In First Out): that means, the first in is the first out. Then, if an assertion of the form of a concept and an assertion of the form of a role are inconsistent, we remove the assertion of the form of a role and keep the assertion of the form of a concept. Otherwise, if two assertions of the form of a concept are inconsistent, we keep the first and delete the second.

Algorithm 1 Most-Possible Repair proposed Algorithm

```
Data: K= ⟨T, A⟩ with A = {S_1, ..., S_n}
Result: A set of consistent assertions M-PR(A)
M-PR(A) ← ∅
for i = 1 to n do
    posmax ← i
    for j = i + 1 to n do
        if S_j > Sposmax then
            posmax ← j
        end
    end
    temp ← Sposmax
    Sposmax ← S_i
    S_i ← temp
end
for i = 1 to n do
    if A role and a concept are inconsistent then
        Delete (role)
    else if Two concepts C1 and C2 are inconsistent then
        Delete (C2)
    M-PR(A) ← M-PR(A) ∪ C_i ∪ R_i (C_i ∪ R_i is a set of assertions in concept or role form)
end
return M-PR(A)
```

4.2 Experimental study and results analysis

The ontology used in this article describes the workflow in a development IT company. Our ontology is encoded by a TBox which contains 43 axioms and an ABox provided by 100 separate sources $\mathcal{A}$ = (S1,..., S100) represented as characteristics of the used ABox data collection : 97 Concept assertions and 56 Role assertions.

The tests are to be applied on ABox which contain *22*, *44* and *66* sets of conflicts respectively. Our algorithm shows a calculation of correct assertions and makes our ontology consistent.

We are interested in the performance indices used to evaluate our algorithm. In our case, we classify ABox assertions into two classes: consistent and inconsistent.

To evaluate our approach, we consider the precision, the recall, and F-measure defined as follows:

The $Precision$ P is calculated by dividing the total number of assertions by the number of consistent assertions. If it's big, it means the method has the most consistent assertions and can be called "precise". The $Recall$ R of the number of consistent assertions found to the total number of consistent assertions is known as the Recall. The $F\text{-}measure$ F is the harmonic average of the precision P and the recall R to give the performance of the system.

$$P = \frac{CE}{CE + IE} \tag{3}$$

$$R = \frac{CE}{CE + CNE} \tag{4}$$

$$F = \frac{2 * (P * R)}{(P + R)} \tag{5}$$

Where CE denotes the number of consistent assertions returned after the algorithms have been applied. CNE is the number of consistent assertions that were not returned after the algorithms were applied. And IE stands for the number of inconsistent assertions that exist before applying the algorithms.

Table 1: **Linear-Based Repair** Algorithm Evaluation Results.

Linear-Based Repair $\ell(\mathcal{A})$			
Conflict size	Precision (P)	Recall (R)	F-measure (F)
22	84,61	77.07	80.66
44	66.41	69.04	67.70
66	48.03	53.04	50.41
Average value	66.35	66.38	66.25

Table 2: **Non-Defeated Repair** Algorithm Evaluation Results.

Non-Defeated Repair nd($\mathcal{A}$)			
Conflict size	Precision (P)	Recall (R)	F-measure (F)
22	84.93	78.98	81.84
44	70.66	84.12	76.81
66	58.75	81.73	68.36
Average value	71.44	81.61	75.67

Table 3: **Most-Possible Repair** Algorithm Evaluation Results.

Most-Possible Repair M-PR($\mathcal{A}$)			
Conflict size	Precision (P)	Recall (R)	F-measure (F)
22	86.16	87.26	86.70
44	73.49	96.82	83.56
66	62.92	97.39	76.45
Average value	74.19	93.82	82.23

By analyzing the results obtained in the series of experiments, there is a difference in efficiency between the various approaches. Indeed, the measurements gathered in tables 1, 2 and 3 indicate that our **Most-Possible Repair** algorithm is the best among the other algorithms. In addition, we find that the Linear-Based Repair and Non-Defeated Repair algorithms do not return a maximum number of consistent assertions. In fact, the repair returned by our algorithm is the largest and most productive than that returned by the Linear-Based Repair and Non-Defeated Repair algorithms. We also noticed that our algorithm is the best compared to the other algorithms.

5 Conclusion and prospects

In this paper, we have proposed a new approach for dealing with inconsistencies in prioritized DL-Lite knowledge bases. To do this, we started by studying the main approaches to dealing with inconsistencies in DL-Lite ontologies existing in the literature. At the end of a synthetic assessment, which we carried out, to compare these approaches according to a certain number of criteria, we were able to propose and detail our algorithm to deal with the inconsistency. We have carried out experiments studies on our approach, using standard performance measures, namely precision, recall and F-measure. The experimental study and analysis have shown that our **Most-Possible Repair** solution is more productive.

As future work, several perspectives remain possible. We cite two of them which are short-term: a first contribution consists in applying our approach to a larger ontology in order to confirm its reliability and robustness. A second possible prospect consists in adding a module to automatically manage the inconsistency when it appears when adding new assertions, through a graphical interface allowing this addition.

References

1. Artale A, Calvanese D, Kontchakov R, Zakharyaschev M, The DL-Lite Family and Relations, CoRR - Computing Research Repository, Volume abs/1401.3487, (2014)
2. Lenzerini M, Ontology-Based Data Management, In: Proceedings of the 6th Alberto Mendelzon International Workshop on Foundations of Data Management, Volume 866, pp. 12–15, ACM, Glasgow, United Kingdom (2011)
3. Hamdi G, Omri M N, Benferhat S, Bouraoui Z, Papini O, Query answering DL-lite knowledge bases from hidden datasets, Annals of Mathematics and Artificial Intelligence, Volume 89, page 271–299 (2021)
4. Hamdi G, Omri M N, Papini O, Benferhat S, Bouraoui Z, Querying DL-lite Knowledge Bases from Hidden Datasets, In: International Symposium on Artificial Intelligence and Mathematics, Fort Lauderdale, Florida, USA (2018)
5. Bienvenu M, Rosati R, Tractable Approximations of Consistent Query Answering for Robust Ontology-based Data Access, In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence, pp. 775–781, IJCAI/AAAI, Beijing, China (2013)
6. Bertossi L E, Database Repairing and Consistent Query Answering, Morgan & Claypool Publishers, (2011)
7. Benferhat S, Dubois D, Prade H, How to infer from inconsistent beliefs without revising?, In: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, pp. 1449–1457, Morgan Kaufmann, Montréal Québec, Canada (1995)
8. Staworko S, Chomicki J, Marcinkowski J, Prioritized repairing and consistent query answering in relational databases, Annals of Mathematics and Artificial Intelligence, Volume 64, page 209–246 (2012)
9. Du J, Qi G, Shen Y, Weight-based consistent query answering over inconsistent over inconsistent SHIQ knowledge base, Knowledge and Information Systems, Volume 34, page 335–371 (2013)
10. Hamdi G, Telli A, Omri M N, Querying of several DL-Lite knowledge bases from various information sources-based polynomial response unification approach, Journal of King Saud University - Computer and Information Sciences, (2020)
11. Telli A, Hamdi G, Omri M N, Lexicographic Repair Under Querying Prioritised DL-Lite Knowledge Bases, Scientific Journal of King Faisal University Basic and Applied Sciences, Volume 22, (2021)
12. Baral C, Kraus S, Minker J, Combining Multiple Knowledge Bases, IEEE Transactions on Knowledge and Data Engineering, Volume 3, page 208–220 (1991)
13. Benferhat S, Bouraoui Z, Tabia K, How to Select One Preferred Assertional-Based Repair from Inconsistent and Prioritized DL-Lite Knowledge Bases?, In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, pp. 1450–1456 (2015)
14. Telli A, Benferhat S, Bourahla M, Bouraoui Z, Tabia K, Polynomial Algorithms for Computing a Single Preferred Assertional-Based Repair, Kunstliche Intelligenz, Volume 31, page 15–30 (2017)
15. Boughammoura R, Omri M N, Querying deep web data bases without accessing to data, In: 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, pp. 597–603, IEEE, Guilin, China (2017)
16. Boughammoura R, Omri M N, Hlaoua L, Information Retrieval from Deep Web Based on Visual Query Interpretation, International Journal of Information Research and Review IJIRR, Volume 2, page 45–59 (2012)
17. Boughammoura R, Hlaoua L, Omri M N, G-Form: A Collaborative Design Approach to Regard Deep Web Form as Galaxy of Concepts, In: 12th International Conference of Cooperative Design, Visualization, and Engineering, pp. 170–174, Springer, Mallorca, Spain (2015)

4. **International Conference on New Trends in Intelligent Software Methodologies, Tools and Techniques (SOMET 2021)**

Article [C4] : New Model for Handling Inconsistency Issues in DL-Lite Knowledge Bases

Auteurs : **Hamdi Ghassen** and Omri Mohamed Nazih

Conférence : The 20th International Conference on New Trends in Intelligent Software Methodologies, Tools and Techniques (**SOMET** 2021), 21-23 September 2021. Cancun, Mexico, pp. 183--196

Classement : **C**

New Model for Handling Inconsistency Issues in DL-Lite Knowledge Bases

Authors	Ghassen Hamdi, Mohamed Nazih Omri
Pages	183 - 196
DOI	10.3233/FAIA210019
Category	Research Article
Series	Frontiers in Artificial Intelligence and Applications
Ebook	Volume 337: New Trends in Intelligent Software Methodologies, Tools and Techniques

Abstract

The lightweight description logic (DL-lite) represents one of the most important logic specially dedicated to applications that handle large volumes of data. Managing inconsistency issues, in order to effectively query inconsistent DL-Lite knowledge bases, is a topical issue. Since assertions (ABoxes) come from a variety of sources with varying degrees of reliability, there is confusion in hierarchical knowledge bases. As a consequence, the inclusion of new axioms is a main factor that causes inconsistency in this type of knowledge base. Often, it is too expensive to manually verify and validate all assertions. In this article, we study the problem of inconsistencies in the DL-Lite family and we propose a new algorithm to resolve the inconsistencies in prioritized knowledge bases. We carried out an experimental study to analyze and compare the results obtained by our proposed algorithm, in the framework of this work, and the main algorithms studied in the literature. The results obtained show that our algorithm is more productive than the others, compared to standard performance measures, namely precision, recall and F-measure.

11/02/2024 23:21 Your Submission to SOMET2021 - hamdi.ghassan@gmail.com - Gmail

Gmail somet 2021

99+
Mail
1
Chat
Meet

Nouveau message

Boîte de récep... 2 080
Messages suivis
En attente
Important
Messages envoyés
Brouillons 120
Catégories
Plus

Libellés

Personnel
Unwanted
Plus

Your Submission to SOMET2021

Boîte de réception ×

Microsoft CMT <email@msr-cmt.org>
À Ghassen

Traduire en français

Dear Ghassen Hamdi

Thank you very much for your submission to The 20th International Conference on Intelligent Software Met
https://atenea.esimecu.ipn.mx/ (SOMET 2021)

The review process for SOMET 2021 has been finalized. Regarding the requirement of the publisher, we hav
quality to the proceedings published by IOS Press.

We wish to inform you that on the basis of the evaluation made by the Program Committee, your paper:

Paper ID: 14
Title : New Model for Handling Inconsistency Issues in DL-Lite Knowledge Bases

HAS BEEN ACCEPTED as a Full paper, (IOS format up to 14 pages).

We would like to congratulate you on the achievement since many high-quality papers have not been select

Please revise the paper based on the reviewer's comments. For getting the reviews to your paper please l
https://cmt3.research.microsoft.com/SOMET2021/Submission/Index
using the Username and Password you used for the original submission.

The paper needs major revision taken carefully all referred comments into account.

ICORE Conference Portal

CORE rankings is now an international collaboration named ICORE

ICORE rankings page | Frequently asked questions

CORE Dora discussion document | Description of Conference Ranks | Guidelines for reporting rankings | Usage of CORE rankings | 2021 Rankings Process | Data used in Rankings and responses to common concerns

somet Search by: All Source: CORE2021

Search

Showing results 1 - 1 of 1

Export

CORE2021 Summary:
A* - 7.45% of 792 ranked venues
A - 16.41% of 792 ranked venues
B - 36.74% of 792 ranked venues
Australasian B - 1.39% of 792 ranked venues
C - 36.24% of 792 ranked venues
Australasian C - 1.77% of 792 ranked venues
Other - 164 total

Title	Acronym	Source	Rank	Note	DBLP	Primary FoR	Comments	Average Rating
International Conference on Intelligent Software Methodologies, Tools, and Techniques (was International Conference on Software Methods and Tools)	SoMeT	CORE2021	C	none	view	4612	0	N/A

May 2021

New Model for Handling Inconsistency Issues in DL-Lite Knowledge Bases

Ghassen HAMDI [a,1], Mohamed Nazih OMRI [a]
[a] *MARS Research Laboratory, University of Sousse, Tunisia*

Abstract. The lightweight description logic (DL-lite) represents one of the most important logic specially dedicated to applications that handle large volumes of data. Managing inconsistency issues, in order to effectively query inconsistent DL-Lite knowledge bases, is a topical issue. Since assertions (ABoxes) come from a variety of sources with varying degrees of reliability, there is confusion in hierarchical knowledge bases. As a consequence, the inclusion of new axioms is a main factor that causes inconsistency in this type of knowledge base. Often, it is too expensive to manually verify and validate all assertions. In this article, we study the problem of inconsistencies in the DL-Lite family and we propose a new algorithm to resolve the inconsistencies in prioritized knowledge bases. We carried out an experimental study to analyze and compare the results obtained by our proposed algorithm, in the framework of this work, and the main algorithms studied in the literature. The results obtained show that our algorithm is more productive than the others, compared to standard performance measures, namely precision, recall and F-measure.

Keywords. DL-Lite, Ontology, Inconsistency, Prioritized Knowledge Bases

1. Introduction

Description logic (DL) is made up of two words: description to represent knowledge and logic to reason from this knowledge. Indeed, DLs are a family of formalisms designed, on the one hand, to represent the knowledge of a domain and on the other hand to reason by deriving new knowledge. Only unary and binary predicates, called concepts and roles, are used in a DL language. To reflect generic knowledge, DLs use only concepts and role inclusions in logical formulas (called axioms) [1].

As we have already mentioned previously, the DL knowledge base (KB) is divided into two parts: a terminological base (called TBox), considered ”generic” or ”global”, true in all models and for all individuals, and an assertional base (called ABox), this information is ”specific” or ”local”, true for certain particular individuals.

DL-Lite [2] is a family of tractable DLs designed for applications that deal with large quantities of data and where the most relevant reasoning function is to respond to queries. DL-Lite ensures a low level of computational complexity. Because of this, DL-Lite is especially well suited to Ontology-Based Data Access (OBDA) [3].

[1] Corresponding Author: Ghassen Hamdi, Mohamed Nazih Omri,; E-mail: hamdighassan@gmail.com, mohamednazih.omri@eniso.u-sousse.tn.

How to deal with inconsistency is a crucially important question that emerges in OBDA ([4,5]. With respect to some assertions that contradict the terminology, inconsistency is defined in such a setting. Usually, a TBox is typically checked and confirmed, while the assertions may be supplied by different and inaccurate sources in vast amounts and can contradict the TBox. Furthermore, manually checking and validating all of the assertions is often prohibitively expensive. This is why, in OBDA, reasoning in the face of inconsistency is important. Several works (*e.g.*[6,7,8,9]), inspired fundamentally by database approaches (*e.g.*[8,9]) or propositional logic approaches (*e.g.*[10,11]), treat with inconsistency in DLs [7], by adapting many inconsistency tolerant inference techniques, named semantics. These are focused on the assertional (or ABox) repair principle [12], which is closely related to the maximally consistent subset concept used in propositional logic [13]. A repair of an ABox is merely a maximum assertional subbase that is consistent with a given TBox [12].

The key issue discussed in this paper is how one chosen repair can be picked. It is necessary to choose only one preferred repair since it allows an accurate query to be answered once the repair is computed. In this paper, the key current inconsistency-tolerant reasoning methods for prioritized KBs are first discussed. Interestingly enough, our approach is appropriate for the DL-Lite environment in the sense that, by producing a single preferred assertional repair, they allow effective handling of inconsistency.

The remainder of this article is arranged in the following manner. The key works are discussed in Section 2 who tried to solve the problem of repairing Inconsistencies in Dl-Lites Knowledge Bases. Section 3 will be dedicated to the presentation of DL-Lite ontology and management of inconsistencies. Section 4 will detail our approach to dealing with inconsistency in a priority knowledge base and the algorithm we have developed for this task. Section 5 provides the experimental study conducted in this work and analyse the obtained results. Section 6 presents a general discussion to give a clear idea on the proposed solution compared to those of the literature. In the final section 7, we present a summary of the work as well as some ideas for future projects.

2. Related Works

In our article, we were inspired by the research area concerning the treatment of inconsistency. Inconsistency is described concerning some assertions that contradict the terminology. Several works (*e.g.* [12,6]) attempted to deal with DL-Lite inconsistency by adapting several inference methods that accept inconsistency, based on database approaches (*e.g.* [9]).

In several applications, assertions come from a variety of sources, some of which may be inconsistent and have varying levels of reliability. Furthermore, different confidence levels may result in different sets of uncertain assertions from the same source. Collecting such sets of assertions provides a prioritized or stratified assertional basis. Priorities play an important role in the management of inconsistency, and they have been extensively discussed in the literature in the sense of propositional logic (*e.g.*[14]). In querying inconsistent databases or DL KBs, numerous works examined the notion of priority (*e.g.*[15,16]). Regretfully, only a few works are available in the OBDA format, such as the one given in [17,18,19] for dealing with reasoning under the prioritized DL-Lite ABox.

In a recent line of research, the inconsistency in lightweight ontology is being investigated. In particular, the writers [7,12,6], look into the issue of KB inconsistency by means of computation a collection of consistent subsets of assertions known as repairs, which recover ontology consistency, and then using them to address queries In addition, in [21,22], polynomials algorithms are proposed by the authors to pick a single preferred repair from a prioritized inconsistent DL-Lite KB, allowing for efficient query answering once the repair has been decided. Specifically, the writers in [21] propose a new approach focused on a single preferred repair solution. However, sequential inference strategies based on the selection of a single coherent assertional basis are proposed by the authors in [22]. In [23,24,25], the authors propose a new algorithm for answering queries without requiring access to Web databases.

In this work, we propose a new approach for dealing with inconsistencies in prioritized DL-Lite knowledge bases in which we will use the principle of cardinality, *i.e.,*the source with the greatest number of assertions has the highest priority. We consider $\mathscr{K} = \langle \mathscr{T}, A \rangle$ a *DL-Lite* knowledge base with priority and we propose a new algorithm called Most-Possible Repair.

3. DL-Lite ontology and management of inconsistencies: an overview

3.1. DL-Lite: A Quick Reminder

Let N_C, N_R and N_I Let NC, NR, and NI represent three pairs of atomic concepts, atomic roles, and individuals that are disjoint. Let $A \in N_C$, $P \in N_R$. Let also '$\neg$', '$\exists$' and '$^-$' three connectors are employed to differentiate complex concepts and roles. The following is a list of *DL-Lite* concepts:

$$B \longrightarrow \quad A \quad | \quad \exists R \qquad C \longrightarrow \quad B \quad | \quad \neg B$$
$$R \longrightarrow \quad P \quad | \quad P^- \qquad E \longrightarrow \quad R \quad | \quad \neg R$$

where A is an atomic concept, P is an atomic role and P^- is the inverse of an atomic role, B (*resp.* C) is the set of basic (*resp.* complex) concepts, and R (*resp.* E) is the set of basic (*resp.*complex) roles.

Let $\mathscr{K}$=$\langle \mathscr{T}, \mathscr{A} \rangle$ be a *DL-Lite* KB where $\mathscr{T}$ stands for TBox, and $\mathscr{A}$ stands for ABox. A TBox is made up of a finite set of inclusion axioms in the form of inclusion assertions on concepts and functions: $B \sqsubseteq C$ (*resp.* negative inclusion assertions $B \sqsubseteq \neg C$) suggests that concept B is a subset of concept C. (*resp.* concept B is not included in the concept C) and $R \sqsubseteq E$ (*resp.* $R \sqsubseteq \neg E$) implying role R is part of role E (*resp.* role R is not included in role E).

The ABox is made up of a limited number of assertions (facts) of the kind $A(a)$ and $P(a;b)$, where $A \in N_C$, $P \in N_R$ and $a, b \in N_I$. The semantics is expressed as interpretations $\mathscr{I}$=$(\Delta^{\mathscr{I}}, .^{\mathscr{I}})$ which is made up of a domain that isn't empty $\Delta^{\mathscr{I}}$ and an interpretation function $.^{\mathscr{I}}$ that allocates to each $a \in N_I$ an element $a^{\mathscr{I}} \in \Delta^{\mathscr{I}}$, to each $A \in N_C$ a subset $A^{\mathscr{I}} \subseteq \Delta^{\mathscr{I}}$ and to each $P \in N_R$ an $P^{\mathscr{I}} \subseteq \Delta^{\mathscr{I}} \times \Delta^{\mathscr{I}}$. If there is a concept C, a TBox $\mathscr{T}$ is said to be incoherent s.t $\forall \mathscr{I}$:$\mathscr{I} \models \mathscr{T}$, we have $C^{\mathscr{I}}$=$\emptyset$. A *DL-Lite* KB $\mathscr{K}$ is alleged to be inconsistent if it refuses to consider any sort of model.

3.2. Inconsistency Treatment for Prioritized DL-Lite Assertional Bases

Data is provided by many potentially contradictory sources with varying levels of reliability in many applications. Indeed, an ontology is usually confirmed and checked, while assertions can be obtained in large amounts from a variety of sources, some of which are unreliable and may be inconsistent with the ontology. It is therefore difficult to create a coherent ontology. This inconsistency can occur for several reasons, such as error modeling, ontology migration or merging and its evolution. Furthermore, manually verifying and validating all assertions is also prohibitively costly.

It is therefore essential to study how to deal with inconsistent ontologies and propose reliable solutions to this problem. From this research perspective, many approaches have been proposed to solve this problem. These approaches are primarily used to deal with inconsistencies in Expressive Description Logics (DL). In our work, we consider the management of inconsistencies in the DL-Lite family. In what follows, we will address the problem of managing inconsistencies in light ontologies when all assertions (facts) are classified as priorities. A priority knowledge base is a set of ABoxes that can be derived from a variety of priority data sources.

In fact, a DL-Lite knowledge base with priority is the set of assertions denoted by: $A = \{S_1, ..., S_n\}$. The sets S_i are called layers and each layer S_i contains the set of assertions with the same priority level i and are considered to be the most reliable than those present in a layer S_j when j¿ i. As a consequence, S_1 includes the most relevant assertions, while S_n contains the less important assertions. In the following, we use the DL-Lite notation $\mathscr{K}$=$\langle \mathscr{T}, \mathscr{A} \rangle$ to link to a prioritized DL-Lite knowledge base of the kind: $A = \{S_1, ..., S_n\}$.

3.3. Example of ontology

Before presenting the approaches cited in the literature to deal with the problem of inconsistency, we cite this example of ontology. We are going to apply all the algorithms that we are going to study on this example.

Our example is defined as: ”Doctor”, ”Internal” and ”HeadOfDepartment” three concepts. Also ”Supervises” and ”HasSupervisor” two roles. Suppose we have the following ontology:

- ”Doctor” are not ”Internal”.
- ”Doctor” must supervise at least one ”Internal”.
- People who are supervised by a ”Doctor” are ”Internal”.
- Each ”Internal” has at least one ”Supervisor”.
- ”Supervisor” is a ”Doctor”.
- ”HeadOfDepartment” is a ”Doctor”.

This ontology is encoded by the following TBox:

$$\boxed{\begin{array}{l} Doctor \sqsubseteq \neg Internal \\ Doctor \sqsubseteq \exists Supervises \\ \exists Supervises^{-} \sqsubseteq Internal \\ Internal \sqsubseteq \exists HasSupervisor \\ \exists HasSupervisor^{-} \sqsubseteq Doctor \\ HeadOfDepartment \sqsubseteq Doctor \end{array}}$$

In addition to this TBox, suppose we have six individuals: Bill, John, Mary, Joe, Bob and Anne. We assume that there is some uncertainty about assertions of our ontology. The available factual information is encoded by the following ABox, which is assumed to be provided by five separate sources $A = \{S_1, S_2, S_3, S_4, S_5\}$ such as:

$$\boxed{\begin{array}{l} S_1 = \{Internal(Bill), Doctor(John)\}, \\ S_2 = \{Doctor(Mary), Internal(Bob)\}, \\ S_3 = \{HasSupervisor(Bob, Bill), \\ \qquad HeadOfDepartment(Joe)\}, \\ S_4 = \{Internal(Anne)\}, \\ S_5 = \{Doctor(Bill)\} \end{array}}$$

In this example, The most accurate assertions are found in S_1, while S_5 includes the least trustworthy ones. Indeed, it is simple to show that the knowledge base is inconsistent. For example, from the axioms $Doctor \sqsubseteq \neg Internal$ and $Internal \sqsubseteq \exists HasSupervisor$ in the TBox, we can conclude that the pairs of ABox assertions ($Internal(Bill)$, $Doctor(Bill)$) and ($Internal(Bill)$, $HasSupervisor(Bob, Bill)$) are inconsistent.

In the following, we represent the main strategies to deal with inconsistent DL-Lite knowledge bases when the assertional base is prioritized.

3.4. Possibilistic-Based Repair

Let $\mathscr{K} = \langle \mathscr{T}, A \rangle$ be an inconsistent prioritized $DL-Lite$ knowledge base. The inconsistency degree of K, denoted by, $Inc(K)$, is defined as follows: $Inc(K) = i+1$ if and only if:

- i) $\langle \mathscr{T}, S_1, ..., S_i \rangle$ is consistent and,
- ii) $\langle \mathscr{T}, S_1, ..., S_i + 1 \rangle$ is inconsistent.

The Possibilistic-Based Repair denoted by $\pi(A)$ is composed of assertions with priority levels that are specifically lower than $Inc(K)$, namely $\pi(A) = S_1 \cup \ldots \cup S_{(Inc(K)-1)}$. If K is consistent then we simply let $\pi(A) = A$.

By applying the Possibilistic-Based Repair on the previous example, we get the following result:

$$\boxed{\begin{array}{l} \pi(A) = S_1 \cup S_2, \\ = \{Internal(Bill), Doctor(John), Doctor(Mary), \\ Internal(BoB)\} \end{array}}$$

So we can notice that $\langle \mathscr{T}, S_1 \cup S_2 \rangle$ is consistent, but $\langle \mathscr{T}, S_1 \cup S_2 \cup S_3 \rangle$ is inconsistent. In addition, we find that the Possibilistic-Based Repair stops at the first layer where inconsistency is introduced. Therefore, only assertions with a degree less than or equal to the degree of inconsistency are taken into account when drawing conclusions. However, assertions with priority levels that are strictly higher than the consistency level are simply inhibited even if they are not involved in any conflict.

In order to find a solution to this limitation and provide larger and more productive repairs, a linear approach can be used.

3.5. Linear-Based Repair

Let $\mathscr{K} = \langle \mathscr{T}, A \rangle$ be an inconsistent and priority DL-Lite knowledge base. The linear based repair of A is indicated by: $\ell(A) = \{S'_1, ..., S'_n\}$

- i) if i=1 then

$$S'_1 = \begin{cases} Sl & \text{if } \langle \mathscr{T}, S_l \rangle \text{ is consistent} \\ \emptyset & Otherwise \end{cases} \tag{1}$$

- ii) for i = 2, ..., n

$$S'_i = \begin{cases} S_i & \text{if } \left\langle \mathscr{T}, S'_1 \cup \ldots \cup S'_{(i-1)} \cup S_i \right\rangle \text{ is consistent} \\ \emptyset & Otherwise \end{cases} \tag{2}$$

Clearly,$\ell(A)$ is acquired by dismissing an S_i layer when its facts contradict with the previous layers. Indeed, the subbase $\ell(A)$ is unique and it is consistent with $\mathscr{T}$.

By applying the Linear-Based Repair on our example, we get this result:

$$\boxed{\begin{array}{l} S'_1 = \{Internal(Bill), Doctor(John)\} \\ S'_2 = \{Doctor(Mary), Internal(BoB)\} \\ S'_3 = \emptyset \\ S'_4 = \{Internal(Anne)\} \\ S'_5 = \emptyset \\ \text{So } \ell(A) = S'_1 \cup S'_2 \cup S'_3 \cup S'_4 \cup S'_2 \\ = S'_1 \cup S'_2 \cup S'_4 \end{array}}$$

This result shows that the sub-bases $\langle \mathscr{T}, S_1 \cup S_2 \rangle$ and $\langle \mathscr{T}, S_1 \cup S_2 \cup S_4 \rangle$ are consistent. However, the sub-bases $\langle \mathscr{T}, S_1 \cup S_2 \cup S_3 \rangle$ and $\langle \mathscr{T}, S_1 \cup S_2 \cup S_4 \cup S_5 \rangle$ are inconsistent. Moreover, the ℓ -consequence is more productive than the π -consequence as we can see in the example of Possibilistic-Based Repair.

In the next section, we present a new strategy that selects a single repair as the preferred one.

3.6. Non-Defeated Repair

Another way to achieve a preferred repair is to interractively recover, layer by layer, all of the free elements free (A).

We consider $\mathscr{K} = \langle \mathscr{T}, A \rangle$ a $DL-Lite$ knowledge base with priority. We denote by $free(A)$ the set of assertions relate to A that aren't to blame for any inconsistencies in $\langle \mathscr{T}, A \rangle$.

The Non-Defeated Repair, denoted by $nd(A)$ as a sequence $nd(A) = S'_1 \cup ... \cup S'_n$, such that:

$$\forall i = 1 \ldots n, S'_i = free(S_1 \cup \ldots \cup S_i) where free(S_1 \cup \ldots \cup S_i)$$

Namely, $nd(A) = free(S_1) \cup free(S_1 \cup S_2) \cup \ldots \cup free(S_1 \cup \ldots \cup S_{S_i})$.

The result obtained by applying the Non-Defeated Repair, in our example, is the following:

$free(S1) = \{Internal(Bill), Doctor(John)\}$
$free(S1 \cup S2) = \{Internal(Bill), Doctor(John),$
$Doctor(Mary), Internal(BoB)\}$
$free(S1 \cup S2 \cup S3) = \{Internal(BoB), Doctor(John),$
$Doctor(Mary), HeadOfDepartment(Joe)\}$
$free(S1 \cup \ldots \cup S4) = \{Internal(BoB), Doctor(John),$
$Doctor(Mary), Internal(Anne),$
$HeadOfDepartment(Joe)\}$
$free(S1 \cup \ldots \cup S5) = \{Doctor(John), Doctor(Mary),$
$Internal(Anne), HeadOfDepartment(Joe),$
$Internal(BoB)\}$

So the result obtained is $nd(A) = \{Internal(Bill), Doctor(John), HeadOfDepartment(Joe),$ $Internal(BoB), Doctor(Mary), Internal(Anne)\}$

4. Most-Possible Repair proposed approach

4.1. Most-Possible Repair proposed algorithm

Our algorithm is presented in this way: First, we initialize M-PR(A) on an empty set. Then we sort the layers according to their cardinalities in a decreasing way (according to the layers that contain the largest number of assertions). If two sources have the same cardinality, we apply FIFO (First In First Out): that means, the first in is the first out. Then, if an assertion of the form of a concept and an assertion of the form of a role are inconsistent, we remove the assertion of the form of a role and keep the assertion of the form of a concept. Otherwise, if two assertions of the form of a concept are inconsistent, we keep the first and delete the second.

4.2. Example of application

We apply algorithm 1 to the same example to which we have applied the algorithms studied in the literature. The results obtained are as follows:

$Step1$: *M-PR(A)* $= \{Internal(Bill), Doctor(John)\}$
$Step2$: *M-PR(A)* $= \{Internal(Bill), Doctor(John),$
$Doctor(Mary), Internal(BoB)\}$
$Step3$: *M-PR(A)* $= \{Internal(Bill), Doctor(John),$
$Doctor(Mary), Internal(BoB),$
$HeadOfDepartment(Joe),$
$Internal(Anne)\}$
$Step4$: *M-PR(A)* $= \{Internal(Bill), Doctor(John),$
$Doctor(Mary), Internal(BoB),$
$HeadOfDepartment(Joe),$
$Internal(Anne)\}$

The algorithm we have proposed returns *M-PR(A)* = Internal(Bill), Doctor(John), Doctor(Mary), Internal(BoB), HeadOfDepartment(Joe), Internal(Anne).

Algorithm 1 Most-Possible Repair proposed Algorithm

Data: $\mathscr{K} = \langle \mathscr{T}, A \rangle$ with $A = \{S_1, ..., S_n\}$
Result: A set of consistent assertions *M-PR(A)*
M-PR(A) $\leftarrow \emptyset$
for $i = 1$ *to* n **do**
  $posmax \leftarrow i$
  for $j = i+1$ *to* n **do**
    if $S_j > Sposmax$ **then**
      $posmax \leftarrow j$
    end
  end
  $temp \leftarrow Sposmax$
  $Sposmax \leftarrow S_i$
  $S_i \leftarrow temp$
end
for $i = 1$ *to* n **do**
  if *A role and a concept are inconsistent* **then**
    Delete (role)
  else if *Two concepts* $C1$ *and* $C2$ *are inconsistent* **then**
    Delete ($C2$)
  M-PR(A) $\leftarrow$ *M-PR(A)* $\cup C_i \cup R_i$ ($C_i \cup R_i$ is a set of assertions in concept or role form)
end
return *M-PR(A)*

This following table summarizes the result obtained by applying the algorithms from the literature and our algorithm to the some example.

Algorithm	Result
Possibilistic-Based Repair	$\pi(A) = \{Internal(Bill), Doctor(John), Doctor(Mary), Internal(BoB)\}$
Linear-Based Repair	$\ell(A) = \{Internal(Bill), Doctor(John) Doctor(Mary), Internal(BoB), Internal(Anne)\}$
Non-Defeated Repair	$nd(A) = \{Internal(Bill), Doctor(John) HeadOfDepartment(Joe), Internal(BoB), Doctor(Mary), Internal(Anne)\}$
Most-Possible Repair	$nd(A) = \{Internal(Bill), Doctor(John) HeadOfDepartment(Joe), Internal(BoB), Doctor(Mary), Internal(Anne)\}$

Table 1.: Comparative table between inconsistency management algorithms.

For the example we are considering, the result returned by the Non-Defeated Repair algorithm and that returned by our Most-Possible Repair algorithm is the same: *nd (A) = M-PR (A)*. Our approach gives a more productive repair than that of the Possibilistic-Based Repair and Linear-Based Repair algorithms.

5. Experimental study and results analysis

5.1. Description of the test corpora

The ontology used in this article describes the workflow in a development IT company. Our ontology is encoded by a TBox which contains 43 axioms and an ABox provided by

100 separate sources A = (S1,..., S100) represented as follows: (i) 97 Concept assertions and (ii) 56 Role assertions.

The tests are to be applied on ABox which contain *22*, *44* and *66* sets of conflicts respectively. Our algorithm shows a calculation of correct assertions and makes our ontology consistent.

Number of conflicts in ABox	Size considered
Set 1	22
Set 2	44
Set 3	66

Table 2.: Characteristics of the used data tests collection

We are interested in the performance indices used to evaluate our algorithm. In our case, we classify ABox assertions into two classes: consistent and inconsistent.

5.2. Evaluation metrics

After that, we cite the precision, recall, F-measure, Accuracy and G-mean, which we'll use to determine the effectiveness of our algorithms. These measures are defined based on the next confusion matrix.

	Expected assertions	
Type of assertion	consistent	inconsistent
Consistent assertions retrieved	CE	INR
Inconsistent assertions	CNE	IE

Table 3. Diagram of confusion matrix

The four efficiency indexes are as follows: *CE* denotes the number of consistent assertions returned after the algorithms have been applied. *CNE* is the number of consistent assertions that were not returned after the algorithms were applied. *IE* stands for the number of inconsistent assertions that exist before applying the algorithms.

5.2.1. The Precision (P)

The Precision is calculated by dividing the total number of assertions by the number of consistent assertions. If it's big, it means the method has the most consistent assertions and can be called "precise".

$$P = \frac{CE}{CE + IE} \tag{3}$$

5.2.2. The Recall (R)

The ratio of the number of consistent assertions found to the total number of consistent assertions is known as the Recall.

$$R = \frac{CE}{CE + CNE} \tag{4}$$

5.2.3. The F-measure (F)

The F-measure is the harmonic average of the precision P and the recall R to give the performance of the system.:

$$F = \frac{2 * (P * R)}{(P + R)} \tag{5}$$

5.3. Experimental evaluations and results

The Tables 5, 6, 7 and 8 show the precision, recall and F-measure, after applying the algorithms (Possibilistic-Based Repair, Linear-Based Repair, Non-Defeated Repair and our Proposed Algorithm Most-Based Repair) on ABoxes which contain 22, 44 and 66 conflict sets respectively.

Table 4. Possibilistic-Based Repair Algorithm Results.

Possibilistic-Based Repair $\pi(A)$			
Conflict size	Precision (P)	Recall (R)	F-measure (F)
22	26.66	5.09	8.55
44	15.38	6.34	8.98
66	10.81	6.95	8.46
Average value	17.61	6.12	8.66

Table 5. Linear-Based Repair
Algorithm Results.

Linear-Based Repair $\ell(A)$			
Conflict size	Precision (P)	Recall (R)	F-measure (F)
22	84,61	77.07	80.66
44	66.41	69.04	67.70
66	48.03	53.04	50.41
Average value	66.35	66.38	66.25

Table 6. Non-Defeated Repair
Algorithm Results.

Non-Defeated Repair nd(A)			
Conflict size	Precision (P)	Recall (R)	F-measure (F)
22	84.93	78.98	81.84
44	70.66	84.12	76.81
66	58.75	81.73	68.36
Average value	71.44	81.61	75.67

Table 7. Most-Possible Repair
Algorithm Results.

Most-Possible Repair M-PR(A)			
Conflict size	Precision (P)	Recall (R)	F-measure (F)
22	86.16	87.26	86.70
44	73.49	96.82	83.56
66	62.92	97.39	76.45
Average value	74.19	93.82	82.23

The following figures 1, 2 and 3 show the 3 graphs that illustrate the Precision, Recall and F-measure according to the size of the conflict of the different ABox used in the four applied algorithms.

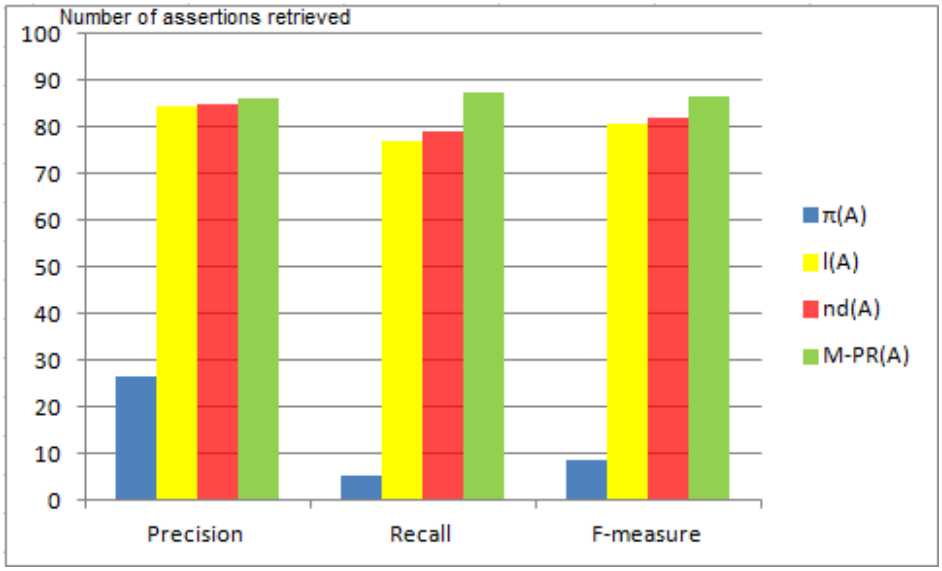


Figure 1.: Evaluation Results for 22 conflicts between different algorithms in term of Precision (P),Recall(R) and F-measure(F).

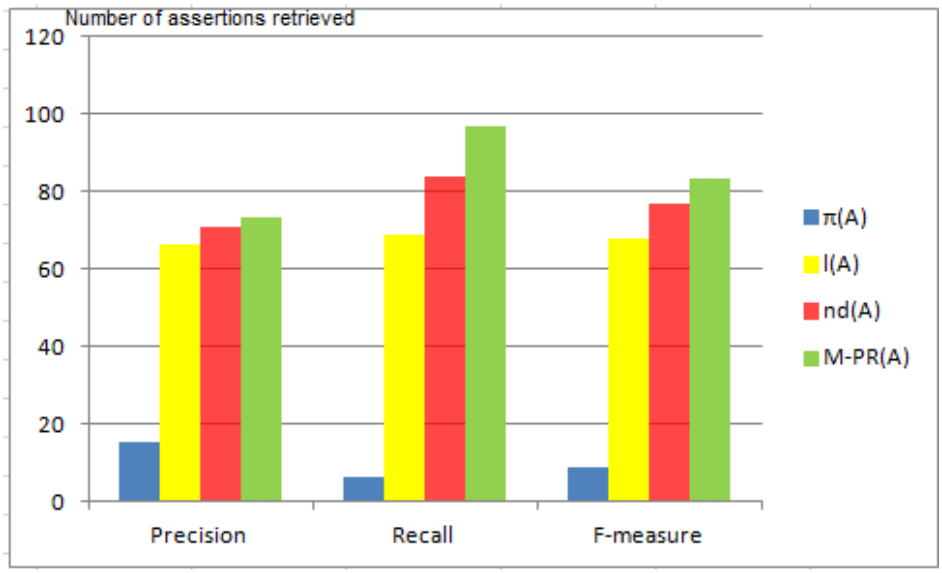


Figure 2.: Evaluation Results for 44 conflicts between different algorithms in term of Precision (P),Recall(R) and F-measure(F).

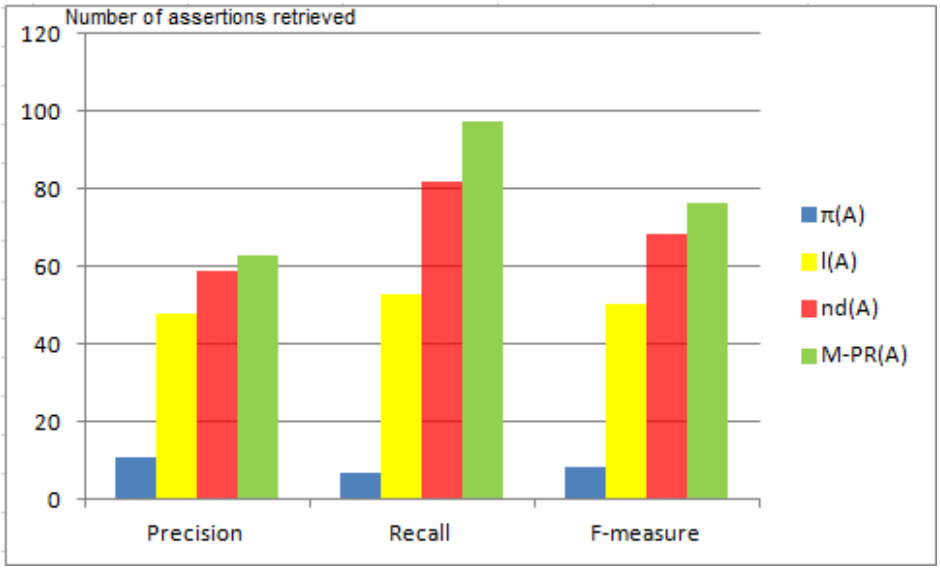


Figure 3.: Evaluation Results for 66 conflicts between different algorithms in term of Precision (P),Recall(R) and F-measure(F).

In the next section, we'll look at how these findings were analyzed and how they compared to other approaches.

6. Discussion

By analyzing the results obtained in the series of experiments, there is a difference in efficiency between the various approaches. Indeed, the measurements gathered in tables 5, 6, 7 and 8 indicate that our **Most-Possible Repair** algorithm is the best among the other algorithms.

In addition, we find that the Possibilistic-Based Repair, Linear-Based Repair and Non-Defeated Repair algorithms do not return a maximum number of consistent assertions. In fact, the repair returned by our algorithm is the largest and most productive than that returned by the Possibilistic-Based Repair, Linear-Based Repair and Non-Defeated Repair algorithms.

Figures 1, 2 and 3 show that:

- As the number of conflicts increases, the precision measurement decreases in the four algorithms. In addition, we notice that our proposed algorithm is more precise than the other algorithms, compared to the different ABoxes.
- With the four algorithms, the recall measure rises as the number of conflicts rises. As a result, the number of conflicts in the ABox has no bearing on this metric.
- As the number of conflicts increases, for each of the four algorithms, the F-measure decreases. As a consequence, the number of conflicts in the ABox has an effect on this metric. As a result, the performance of the four algorithms and especially the Most-Possible Repair is degraded.

Finally, according to the precision measures presented in Tables 5, 6, 7 and 8 and from the previous figures, we also noticed that our **Most-Possible Repair** algorithm is the best algorithm compared to the other algorithms.

All the approaches studied return a consistent assertional basis, but they are influenced by the increase in the number of conflicts in the ABox.

To conclude, our proposed **Most-Possible Repair** approach is the most productive and efficient, while the Possibilistic-Based Repair algorithm is the least productive among all the algorithms studied.

In addition, our proposed **Most-Possible Repair** algorithm gives the best result compared to the other algorithms studied and returns the largest number of consistent assertions.

7. Conclusion and prospects

Ontologies have taken an important place in knowledge-based applications. Today, a growing number of ontologies are widely available on the web for reuse but the addition of a new axiom, in an existing ontology, can generate an inconsistent knowledge base. Managing these inconsistencies in this type of knowledge bases, particularly assertional and prioritized, thus becomes a crucial problem in many applications.

In this context, we have proposed a new approach for dealing with inconsistencies in prioritized DL-Lite knowledge bases. To do this, we started by studying the main approaches to dealing with inconsistencies in DL-Lite ontologies existing in the literature. At the end of a synthetic assessment, which we carried out, to compare these approaches according to a certain number of criteria, we were able to propose and detail our algorithm to deal with the inconsistency. We have carried out experiments studies on our approach, using standard performance measures, namely precision, recall and F-measure. The experimental study and analysis, which we have carried out, obtained results have shown that our **Most-Possible Repair** solution is more productive.

As future work, several perspectives remain possible. We cite two of them which are short-term: a first contribution consists in applying our approach to a larger ontology in order to confirm its reliability and robustness. And, a second possible prospect which consists in adding a module to automatically manage the inconsistency when it appears when adding new assertions, through a graphical interface allowing this addition.

References

[1] Baader F, Calvanese D, McGuinness D, Daniele N, Nardi D, Patel-Schneider P F, The description logic handbook: theory, implementation, and applications, Cambridge University Press, New York, USA (2003)

[2] Artale A, Calvanese D, Kontchakov R, Zakharyaschev M, The DL-Lite Family and Relations, CoRR - Computing Research Repository, Volume abs/1401.3487, (2014)

[3] Rodriguez-Muro M, Kontchakov R, Zakharyaschev M, Ontology-Based Data Access: Ontop of Databases, In: 12th International Semantic Web Conference, Volume 8218, pp. 558–573, Springer, Sydney, NSW, Australia (2013)

[4] Hamdi G, Omri M N, Benferhat S, Bouraoui Z, Papini O, Query answering DL-lite knowledge bases from hidden datasets, Annals of Mathematics and Artificial Intelligence, Volume 89, page 271–299 (2021)

[5] Hamdi G, Omri M N, Papini O, Benferhat S, Bouraoui Z, Querying DL-lite Knowledge Bases from Hidden Datasets, In: International Symposium on Artificial Intelligence and Mathematics, Fort Lauderdale, Florida, USA (2018)

[6] Bienvenu M, Rosati R, Tractable Approximations of Consistent Query Answering for Robust Ontology-based Data Access, In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence, pp. 775–781, IJCAI/AAAI, Beijing, China (2013)

[7] Lembo D, Lenzerini M, Rosati R, Ruzzi M, Savo D F, Inconsistency-tolerant query answering in ontology-based data access, Journal of Web Semantics, Volume 33, page 3–29 (2015)

[8] Arenas M, Bertossi L E, Chomicki J, Consistent Query Answers in Inconsistent Databases, In: Proceedings of the Eighteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, pp. 68–79, ACM Press, Philadelphia, Pennsylvania, USA (1999)

[9] Bertossi L E, Database Repairing and Consistent Query Answering, Morgan & Claypool Publishers, (2011)

[10] Benferhat S, Cayrol C, Dubois D, Lang J, Prade H, Inconsistency Management and Prioritized Syntax-Based Entailment, In: Proceedings of the 13th International Joint Conference on Artificial Intelligence, pp. 640–647, Morgan Kaufmann, Chambery, France (1993)

[11] Benferhat S, Dubois D, Prade H, Some Syntactic Approaches to the Handling of Inconsistent Knowledge Bases: A Comparative Study Part 1: The Flat Case, Studia Logica, Volume 58, page 17–45 (1997)

[12] Lembo D, Lenzerini M, Rosati R, Ruzzi M, Savo D F, Inconsistency-Tolerant Semantics for Description Logics, In: Fourth International Conference of Web Reasoning and Rule Systems RR, pp. 103–117, Springer, Bressanone/Brixen, Italy (2010)

[13] Rescher N, Manor R, On inference from inconsistent premisses, Theory and Decision, Volume 1, page 179-217 (1970)

[14] Benferhat S, Dubois D, Prade H, How to infer from inconsistent beliefs without revising?, In: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, pp. 1449–1457, Morgan Kaufmann, Montréal Québec, Canada (1995)

[15] Staworko S, Chomicki J, Marcinkowski J, Prioritized repairing and consistent query answering in relational databases, Annals of Mathematics and Artificial Intelligence, Volume 64, page 209–246 (2012)

[16] Du J, Qi G, Shen Y, Weight-based consistent query answering over inconsistent over inconsistent SHIQ knowledge base, Knowledge and Information Systems, Volume 34, page 335–371 (2013)

[17] Bienvenu M, Bourgaux C, Goasdoué F, Querying inconsistent description logic knowledge bases under preferred repair semantics, In: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, pp. 996–1002, AAAI Press, Québec City, Québec, Canada (2014)

[18] Hamdi G, Telli A, Omri M N, Querying of several DL-Lite knowledge bases from various information sources-based polynomial response unification approach, Journal of King Saud University - Computer and Information Sciences, (2020)

[19] Telli A, Hamdi G, Omri M N, Lexicographic Repair Under Querying Prioritised DL-Lite Knowledge Bases, Scientific Journal of King Faisal University Basic and Applied Sciences, Volume 22, (2021)

[20] Baral C, Kraus S, Minker J, Combining Multiple Knowledge Bases, IEEE Transactions on Knowledge and Data Engineering, Volume 3, page 208–220 (1991)

[21] Benferhat S, Bouraoui Z, Tabia K, How to Select One Preferred Assertional-Based Repair from Inconsistent and Prioritized DL-Lite Knowledge Bases?, In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, pp. 1450–1456 (2015)

[22] Telli A, Benferhat S, Bourahla M, Bouraoui Z, Tabia K, Polynomial Algorithms for Computing a Single Preferred Assertional-Based Repair, Kunstliche Intelligenz, Volume 31, page 15–30 (2017)

[23] Boughammoura R, Omri M N, Querying deep web data bases without accessing to data, In: 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, pp. 597–603, IEEE, Guilin, China (2017)

[24] Boughammoura R, Omri M N, Hlaoua L, Information Retrieval from Deep Web Based on Visual Query Interpretation, International Journal of Information Research and Review IJIRR, Volume 2, page 45–59 (2012)

[25] Boughammoura R, Hlaoua L, Omri M N, G-Form: A Collaborative Design Approach to Regard Deep Web Form as Galaxy of Concepts, In: 12th International Conference of Cooperative Design, Visualization, and Engineering, pp. 170–174, Springer, Mallorca, Spain (2015)

5. **International Symposium on Artificial Intelligence and Mathematics (ISAIM 2018)**

Article [C5] : Querying DL-lite Knowledge Bases from Hidden Datasets

Auteurs : **Hamdi Ghassen,** Omri Mohamed Nazih, Papini Odile, Benferhat Salem and Zied Bouraoui

Conférence : International Symposium on Artificial Intelligence and Mathematics (**ISAIM** 2018), January 3-5, 2018. Fort Lauderdale, Florida, USA

Classement : **B**

Communication Dans Un Congrès **Année : 2018**

Querying DL-lite Knowledge Bases from Hidden Datasets *

Ghassen Hamdi , Mohamed Nazih Omri , Odile Papini (1, 2) , Salem Benferhat (3) , Zied Bouraoui (3)

Afficher plus de détails

1 LIS - Laboratoire d'Informatique et des Systèmes
2 LIRICA - Logique, Interaction, Raisonnement et Inférence, Complexité, Algèbre
3 CRIL - Centre de Recherche en Informatique de Lens

11/02/2024 23:33 ISAIM 2018 notification for paper 9 - hamdi.ghassan@gmail.com - Gmail

Gmail

isaim 2018

99+
Mail
1
Chat
Meet

Nouveau message

Boîte de récep... 2 080
Messages suivis
En attente
Important
Messages envoyés
Brouillons 120
Catégories
Plus

Libellés

Personnel
Unwanted
Plus

Traduire en français

Dear *NAME*,

this is to inform you that your paper

TITLE

has been accepted for presentation at ISAIM 2018: Congratulations!

Please do prepare a final version (taking the reviewers comments into account) by November 27, so that preliminary informal proceedings can be compiled.

If your paper has been submitted to a special track, the responsible organizers of that track will contact you.

Please do register for the conference (at the website) and arrange your hotel and travel early. We expect all papers to be presented at the conference.

A special issue of AMAI is planned for selected papers (expanded versions of the submissions). Details will be given at the conference.

We hope to see you all early next year for a successful conference with fruitful discussions.

With our best wishes

Jürgen and Guillermo.

ICORE Conference Portal

CORE rankings is now an international collaboration named ICORE

ICORE rankings page | Frequently asked questions

CORE Dora discussion document | Description of Conference Ranks | Guidelines for reporting rankings | Usage of CORE rankings | 2021 Rankings Process | Data used in Rankings and responses to common concerns

isaim Search by: All Source: CORE2018

Search

Showing results 1 - 1 of 1

Export

CORE2018 Summary:
A* - 4.49% of 1492 ranked venues
A - 14.88% of 1492 ranked venues
B - 27.48% of 1492 ranked venues
Australasian B - 0.0% of 1492 ranked venues
C - 53.15% of 1492 ranked venues
Australasian C - 0.0% of 1492 ranked venues
Other - 133 total

Title	Acronym	Source	Rank	Note	DBLP	Primary FoR	Comments	Average Rating
International Symposium on Artificial Intelligence and Mathematics	ISAIM	CORE2018	B	none	view	0801	0	N/A

HAL
open science

Querying DL-lite Knowledge Bases from Hidden Datasets *

Ghassen Hamdi, Mohamed Nazih Omri, Odile Papini, Salem Benferhat, Zied Bouraoui

To cite this version:

Ghassen Hamdi, Mohamed Nazih Omri, Odile Papini, Salem Benferhat, Zied Bouraoui. Querying DL-lite Knowledge Bases from Hidden Datasets *. International Symposium on Artificial Intelligence and Mathematics, Jan 2018, Fort Lauderdale, United States. hal-02084470

HAL Id: hal-02084470

https://hal.science/hal-02084470

Submitted on 29 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Querying DL-lite Knowledge Bases from Hidden Datasets*

Ghassen Hamdi
MARS Research Laboratory
Univ Sousse, Tunisia
hamdighassan@gmail.com

Mohamed Nazih Omri
MARS Research Laboratory
Univ Sousse, Tunisia
mohamednazih.omri@fsm.rnu.tn

Odile Papini
LSIS, Univ Aix-Marseille, France
odile.papini@amu-univ.fr

Salem Benferhat
CRIL, Univ Artois, France
benferhat@cril.fr

Zied Bouraoui
CRIL, Univ Artois, France
bouraoui@cril.fr

Abstract

One of the main aims of Ontology-Based Data Access is to uniform querying data using knowledge encoded in an ontology. Data are often provided by several information sources, and this has led to a number of methods that merge them in order to get a unified point of view. Existing merging approaches relie on the fact that the content of the datasets is known. However in many applications such as in recommendation sites, this prior knowledge about the content of the datasets is not available. This paper investigates the problem of querying multiple information sources without knowing the content of the datasets. We first provide several strategies to answer queries from datasets. We then study how these strategies can be compared to each other from a productivity point of view.

Introduction

Structured knowledge about entities plays an important role in many web applications. Ontologies offer a powerful framework to encode structured knowledge about the concepts and properties of a given domain. They are typically expressed using description logics (Baader et al. 2010), and stored in two parts: a TBox, which encodes generic knowledge about a domain through the semantic relationships between concepts and relations, and an ABox, which contains data, i.e. information about which entities belong (resp. related) to what concepts (resp. relations).

Recent years have witnessed an increasing interest in Ontology-Based Data Access (Maurizio 2011; Poggi et al. 2008; Muro, Kontchakov, and Zakharyaschev 2013), in which structured knowledge (i.e. stored in a TBox) is exploited when querying data (i.e. stored as an ABox). Adding ontological knowledge aims, for instance, at improving query answering, merging/integrating data from different sources, etc.

In the case, when the data are provided from different sources an important research question is: how to perform query answering in a meaningful way? In a multi sources context, one should take into account of many problems such as the dissatisfaction of the sources, missing data, redundancy, etc. These problems have led to a number of works mainly aiming at merging datasources (i.e. ABoxes) in order to get a global point of view (Benferhat et al. 2014; Wang et al. 2012; Benferhat et al. 2017). To perform merging while dealing with conflicts, existing approaches require to perform some fusion operators which allow for combining the pieces of information coming from several sources (ABoxes) in order to know the whole content of the ABoxes. Unfortunately, this is not always guaranteed.

This paper studies the problem of query answering without neither merging datasources nor knowing their contents. This might make sense in certain practical settings. Consider for example the case of recommendation sites, online booking sites where the results given to users are made by only querying and without knowing the whole contents of the sources. The assumption that the data are not accessible or hidden for privacy or security reasons, is quite common, in many Web applications, however as far as we know this problem has not already been addressed in query answering mediated by an ontology.

In this paper, we provide several strategies to answer queries from hidden datasets and we study how these strategies can be compared to each other from a productivity point of view.

Consider a simple example of a Web site managing airline companies described using a rule saying that a customer can book a flight proposed by a company (for the sake of simplicity, a flight concerns departure, destination and date). Consider the flights booking as statements. Thus, we have a dataset for each company. Let us consider q_1 ("Output the list of customers of all companies") and q_2 ("A customer that booked the same flight in different companies"). For privacy purpose, it is obvious that one may not have response for query q_1. On the other hand, one may have response to q_2, but this requires merging the responses given by datasets.

The rest of this paper is organized as follows. In the next section we recall some basic notions of Description Logics and query answering. Subsequently we propose new strategies for querying multiple hidden sources. Finally, we compare the proposed strategies from a productivity point of view and present related works before concluding.

Description Logics

Description Logics (DL) are the logical framework underlying the ontology language, OWL. DL-Lite (Alessandro et

*This work was supported by the european projects, H2020 Marie Skodowska-Curie Actions (MSCA) Research and Innovation Staff Exchange (Rise) : Aniage (Project Number 69—215)

al. 2009) is a family of tractable DLs specifically tailored for applications that use large amounts of data, for which query answering is the most important reasoning task. DL-Lite guarantees a low computational complexity. This fact makes DL-Lite especially well suited for Ontology-Based Data Access (OBDA). In the following, we will recall $DL-Lite_R$ logic.

DL-Lite syntax and semantics. Let N_C, N_R and N_I be three pairwise disjoint sets of atomic concepts, atomic roles and individuals respectively. Let $A \in N_C$, $P \in N_R$, three connectors '$\neg$', '$\exists$' and '$^-$' are used to define complex concepts and roles. Basic concepts (*resp.* roles) B (*resp.* R) and complex concepts (*resp.* roles) C (*resp.* E) are defined in DL-Lite as follows:

$$B \longrightarrow \quad A \quad | \quad \exists R \qquad C \longrightarrow \quad B \quad | \quad \neg B$$
$$R \longrightarrow \quad P \quad | \quad P^- \qquad E \longrightarrow \quad R \quad | \quad \neg R$$

where P^- represents the inverse of P. A DL-Lite knowledge base (KB) is a pair $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ where $\mathcal{T}$ is called the TBox and $\mathcal{A}$ is called the ABox. A TBox includes a finite set of inclusion axioms on concepts and on roles respectively of the form: $B \sqsubseteq C$ and $R \sqsubseteq E$. The ABox contains a finite set of assertions (facts) of the form $A(a)$ and $P(a, b)$ where $A \in N_C$, $P \in N_R$ and $a, b \in N_I$.

The semantics is given in term of interpretations. An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, .^{\mathcal{I}})$ where $\Delta^{\mathcal{I}}$ is called domain, and $.^{\mathcal{I}}$ an interpretation function that assigns to each $a \in N_I$ an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, to each $A \in N_C$ a subset $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and to each $P \in N_R$ an $P^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. The function $.^{\mathcal{I}}$ is extended in a straightforward way for complex concepts and roles, e.g $(\neg B)^{\mathcal{I}}$=$\Delta^{\mathcal{I}} \setminus B^{\mathcal{I}}$, $(P^-)^{\mathcal{I}}$=$\{(y, x) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid (x, y) \in P^{\mathcal{I}}\}$ and $(\exists R)^{\mathcal{I}}$=$\{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} \mid (x, y) \in R^{\mathcal{I}}\}$. An interpretation $\mathcal{I}$ is said to be a model of a concept (*resp.* role) inclusion axiom, denoted by $\mathcal{I} \models B \sqsubseteq C$ (*resp.* $\mathcal{I} \models R \sqsubseteq E$), iff $B^{\mathcal{I}} \subseteq C^{\mathcal{I}}$ (*resp.* $R^{\mathcal{I}} \subseteq E^{\mathcal{I}}$). Similarly, we say that $\mathcal{I}$ satisfies a concept (*resp.* role) assertion, denoted by $\mathcal{I} \models A(a)$ (*resp.* $I \models P(a, b)$), iff $a^{\mathcal{I}} \in A^{\mathcal{I}}$ (*resp.* $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in P^{\mathcal{I}}$). An interpretation $\mathcal{I}$ is said to satisfy a KB $\mathcal{K}$=$\langle \mathcal{T}, \mathcal{A} \rangle$, denoted $\mathcal{I} \models \mathcal{K}$, iff $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \models \mathcal{A}$. Such interpretation is said to be a model of $\mathcal{K}$. Lastly, a TBox $\mathcal{T}$ is said to be incoherent if there exists a concept C s.t $\forall \mathcal{I}$:$\mathcal{I} \models \mathcal{T}$, we have $C^{\mathcal{I}}$=$\emptyset$. A DL-Lite KB $\mathcal{K}$ is said to be inconsistent if it does not admit any model.

Query answering. A query is a first-order logic formula, denoted q=$\{\vec{x} | \phi(\vec{x})\}$, where $\vec{x}$=(x_1,...,x_n) are free variables, n is the arity of q and atoms of $\phi(\vec{x})$ are of the form $A(t_i)$ or $P(t_i, t_j)$ with $A \in N_C$ and $P \in N_R$ and t_i, t_j are terms, *i.e.,* constants of N_I or variables. When $\phi(\vec{x})$ is of the form $\exists \vec{y}.conj(\vec{x}, \vec{y})$ where $\vec{y}$ are bound variables called existentially quantified variables, and $conj(\vec{x}, \vec{y})$ is a conjunction of atoms of the form $A(t_i)$ or $P(t_i, t_j)$ with $A \in N_C$ and $P \in N_R$ and t_i, t_j are terms, then q is said to be a conjunctive query (CQ). When n=0, then q is called a boolean query (BQ). A BQ with no bound variables is called a ground query (GQ). Lastly, when q only contains one atom with no free variables, then it is called an instance query (IQ) (*i.e.,* instance checking).

An answer to a CQ $q(\vec{x}) \leftarrow conj(\vec{x}, \vec{y})$ over a KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ is a non empty set of tuples $\vec{s} = (s_1, \cdots, s_k) \in N_I \times \cdots \times N_I$ such that $\langle \mathcal{T}, \mathcal{A} \rangle \models q(\vec{s})$ $\langle \mathcal{T}, \mathcal{A} \rangle \models q(\vec{s})$.

An answer to a boolean query BQ $q()$ w.r.t. $\mathcal{A}$ is $true$ if $\langle \mathcal{T}, \mathcal{A} \rangle \models q()$ and false otherwise.

In this paper, we focus on conjonctive queries (CQs).

Multiple Sources Query Answering Strategies

This section investigates querying multiple ABox, without knowing their contents, using a DL-Lite ontology. We first introduce the notion of DL-Lite knowledge base with multiple assertional sets. While a standard DLs knowledge base has a single ABox, it is convenient for the definitions of inference strategies to introduce the notion of "profile" which is simply a multiset of ABoxes. We denote by $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ a knowledge base with profile where $\mathcal{T}$ is a standard DL-Lite TBox and $P = \{\mathcal{A}_1, ..., \mathcal{A}_m\}$ is an ABox profile. We assume that each ABox $A_i \in P$ is consistent with the ontology, namely $\langle \mathcal{T}, \mathcal{A}_i \rangle$ is consistent.

Example 1. *Let $Customer$, $Flight$ be two concepts and $books$ a role. Consider the following knowledge base $\mathcal{K}_P = \langle \mathcal{T}, P = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\} \rangle$ where $\mathcal{T} = \{Customer \sqsubseteq \exists books, Flight \sqsubseteq \exists books^-\}$ and. $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ are respectively the datasets corresponding to "AirFrance", "KLM" and "Transavia". The ABoxes are described as follows: $\mathcal{A}_1$ = {Customer(a), Customer(b), Customer(d), Flight(f1), Flight(f2), books(a, f1), books(b, f1), books(d, f2) }, $\mathcal{A}_2$ = {Customer(a), Customer(b), Flight(f1), books(a, f1), books(b, f1)}, $\mathcal{A}_3$ = {Customer(a), Customer(c), Customer(d), Flight(f1), Flight(f2), books(a, f1), books(c, f2), books(d, f2)}.*

This example is used as a running example in the paper in order to illustrate the different proposed strategies to query a knowledge base without knowing the content of the datasets.

Query Answering Strategies

The following definition generalizes query answering for a multiple set of ABoxes.

Definition 1. *Let $\mathcal{K}_P$=$\langle \mathcal{T}, P \rangle$ be an knowledge base with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$ and $q(\vec{x})$ be a conjunctive query. An answer to the query $q(\vec{x})$ w.r.t. $\mathcal{A}_i$ is a tuple $\vec{s} = (s_1, \cdots, s_k) \in N^I \times \cdots \times N^I$ such that $\langle \mathcal{T}, \mathcal{A}_i \rangle \models q(\vec{s})$.*

Let $q(\vec{x})$ be a conjunctive query, S_i denotes the set of answers to $q(\vec{x})$ w.r.t. $\mathcal{A}_i$. Namely $S_i = \{\vec{s} \in N^I \times \cdots \times N^I : \langle \mathcal{T}, \mathcal{A}_i \rangle \models q(\vec{s})\}$. We denote by $S_P = \{S_1, ..., S_m\}$ the profile, multiset of sets of answers to a query $q(\vec{x})$ with respect to P, with S_i be the set of answers to $q(\vec{x})$ w.r.t. $\mathcal{A}_i$ for $1 \leq i \leq m$. Intuitively, S_P contains all the valid answers to a query, namely the answers that can be obtained from each ABox $\mathcal{A}_i \in P$. Of course, when there is no answer to a query $q(\vec{x})$ with respect to $\mathcal{A}_i$, S_i = $\emptyset$. However having a valid answer does not mean that all the sources are in agreement with that. For instance, even though each ABox is consistent with the TBox, it might be the case that the union of these ABoxes is inconsistent. Moreover it might be the case that the sources do not share same information. In

fact, one needs to access to the data stored in the ABoxes in a meaningful way.

We introduce different strategies for query answering without knowing the content of the dataset in $\mathcal{K}_P$. One way to perform query answering is to only consider answers to the query that hold in all the sources. More formally :

Definition 2. *Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a knowledge base with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$. Let $q(\vec{x})$ be a conjunctive query and S_P be the profile of sets of answers ("answers profile") to $q(\vec{x})$ w.r.t. P. An (*universal*) answer to the query $q(\vec{x})$ w.r.t. $\mathcal{K}_P$ is a tuple $\vec{s}$ such that*

$$\mathcal{K}_P \models_\forall q(\vec{s}) \quad if \quad \forall S_i \in S_P,\ \vec{s} \in S_i.$$

Note that the notion of universal inference offers a natural way to deal with conflicting sources. It was used for instance in default reasoning (Reiter 1987), where one only accepts conclusions derived from each extension of a default theory. We denote by $S_\forall$ the set of answers obtained by universal inference, $S_\forall = \{\vec{s} : \mathcal{K}_P \models_\forall q(\vec{s})\}$.

Example 2. *Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be the DL-Lite knowledge base (KB) managing flight booking introduced in Example 1. Assume that an authorized query is a query which asks for the clients that made a booking for a flight, more formally, $q(x) = \exists x\, books(x, y)$. We get the following answers $S_P = \{S_1, S_2, S_3\}$ with $S_1 = \{a, b, d\}$ the set of answers from the ABox $\mathcal{A}_1$, $S_2 = \{a, b\}$ the set of answers from the ABox $\mathcal{A}_2$ and $S_3 = \{a, c, d\}$ the set of answers from the ABox $\mathcal{A}_3$. Clearly, $\mathcal{K}_P \models_\forall q(a)$ since $\forall S_i \in S_P\ a \in S_i$ and $S_\forall = \{a\}$ which expresses the fact that the customer a booked a flight in all airline companies.*

In some situations, it makes sense to only looking for a possible solution of a set of constraints or preferences such as in some decision problem. As a second strategy we propose the existential inference that states that an answer is deduced from at least one ABox. More formally.

Definition 3. *Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a knowledge base with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$. Let $q(\vec{x})$ be a conjunctive query and S_P be the answers profile to $q(\vec{x})$ w.r.t. P. An (*existential*) answer to the query $q(\vec{x})$ w.r.t. $\mathcal{K}_P$ is a tuple $\vec{s}$ such that*

$$\mathcal{K}_P \models_\exists q(\vec{s}) \quad if \quad \exists S_i \in S_P,\ \vec{s} \in S_i$$

We denote by $S_\exists$ the set of answers obtained by existential inference, $S_\exists = \{\vec{s} : \mathcal{K}_P \models_\exists q(\vec{s})\}$.

Example 3. *Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be the KB from Example 2. Consider again the same query $q(x) = \exists x\, books(x, y)$. Clearly, $\mathcal{K}_P \models_\exists q(a)$ since $a \in S_1$ moreover, $\mathcal{K}_P \models_\exists q(b)$, $\mathcal{K}_P \models_\exists q(c)$ and $\mathcal{K}_P \models_\exists q(d)$, therefore $S_\exists = \{a, b, c, d\}$ which expresses the fact that the customers a, b, c and d booked a flight in at least one airline company.*

The existential query answering strategy is a very adventurous strategy. To this end, we propose to consider a majority-based strategy as an alternative. It states that an answer to a query is considered as valid if it can be deduced from the majority of the datasets. More formally,

Definition 4. *Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a knowledge base with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$. Let $q(\vec{x})$ be a conjunctive query and S_P be answers profile to $q(\vec{x})$ w.r.t. P. An (*majority*) answer to the query $q(\vec{x})$ w.r.t. $\mathcal{K}_P$ is a tuple $\vec{s}$ such that*

$$\mathcal{K}_P \models_{maj} q(\vec{s}) \quad if \quad \frac{\mid i\ :\ \vec{s} \in S_i \mid}{m} > \frac{1}{2}$$

We denote by S_{maj} the set of answers obtained by majority-based strategy, namely $S_{maj} = \{\vec{s} : \mathcal{K}_P \models_{maj} q(\vec{s})\}$. Note that this strategy offers a good compromise between universal or existential strategies.

Example 4. *Let us continue Example 2 with the same query : $q(x) = \exists x\, books(x, y)$. One can check that $\mathcal{K}_P \models_{maj} q(b)$ since $b \in S_1$ and $b \in S_2$, $\mathcal{K}_P \models_{maj} q(d)$ since $d \in S_1$ and $d \in S_3$ and $\mathcal{K}_P \models_{maj} q(a)$ since $a \in S_1$, $a \in S_2$, and $a \in S_3$, therefore $S_{maj} = \{a, b, d\}$ which expresses the fact that the customers a, b and d booked a flight in the majority of the airline companies.*

Introducing Cardinality

We now introduce new strategies obtained by adding a cardinality criterion to the strategies proposed above. We basically restrict the application of the query answering relations $\forall, \exists$ and maj to some sets of answers in the profile S_p. More precisely, we only select the largest sets in S_P with respect to cardinality . The choice of the largest sets of answers might make sense in certain practical applications. For instance, suppose that we want to know the customers that booked a flight in the most popular airline compagnies in terms of number of customers.

We denote by Sc_P the profile of sets of answers with maximal cardinality to a query $q(\vec{x})$ w.r.t. P. The set Sc_P is defined as follows:

$$Sc_P = \{S_i,\ i \in \{1, \cdots, m\}\ :\ \forall j \in \{1, \cdots, m\}\ \mid S_j \mid \leq \mid S_i \mid\}$$

The strategies presented before can be extended using cardinality criterion as follows:

Definition 5. *Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a knowledge base with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$. Let $q(\vec{x})$ be a conjunctive query and Sc_P be the answers profile with maximal cardinality to $q(\vec{x})$ w.r.t. P. An answer to the query $q(\vec{x})$ w.r.t. $\mathcal{K}_P$ is a tuple $\vec{s}$ such that :*

- $\mathcal{K}_P \models_{\forall_c} q(\vec{s})$ *if* $\forall S_i \in Sc_P, \vec{s} \in S_i$.
- $\mathcal{K}_P \models_{\exists_c} q(\vec{s})$ *if* $\exists S_i \in Sc_P, \vec{s} \in S_i$.
- $\mathcal{K}_P \models_{maj_c} q(\vec{s})$ *if* $\frac{|S_i \in Sc_P\ :\ \vec{s} \in S_i|}{|Sc_P|} \geq \frac{1}{2}$.

We denote by $S_{\forall_c}$ (resp. $S_{\exists_c}$, S_{maj_c}) the set of answers obtained by universal (resp. existential, majority-based) strategies.

Example 5. *Let us continue Example 2. Clearly, $Sc_P = \{S_1, S_3\}$ then $S_{\forall_c} = \{a, d\}$ which expresses the fact that the customers a and d booked a flight in all the most popular airline companies. $Sc_P = \{S_1, S_3\}$ then $S_{\exists_c} = \{a, b, c, d\}$ which expresses the fact that the customers a, b, c and d booked a flight in at least one of the most popular airline companies.*

Example 6. *Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ with $P = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4\}$ where $\mathcal{T}$ and $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$, are from example 2 and the ABox $\mathcal{A}_4$ is the dataset corresponding to "TunisAir". Consider*

the same query $q(x) = \exists x\, books(x, y)$ and the set of answers to the query q from $\mathcal{A}_4$: S_4 = {a, c, e}. Clearly, $Sc_P = \{S_1, S_3, S_4\}$ and $S_{maj_c} = \{a, c, d\}$ which expresses the fact that the customers a, c and d booked a flight in the majority of the most popular airline companies.

Productivity of Query Answering Strategies

In this section, we compare the proposed query answering strategies from a productivity point of view. We formalize this comparison as follows:

Definition 6. *Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a knowledge base and let s_i and s_j be two strategies. We say that s_i is more productive than s_j, if for any ABox profile $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$ and any query $q(\vec{x})$, the set of answers to $q(\vec{x})$ w.r.t. P using the strategy s_j is included or equal to the set of answers to $q(\vec{x})$ w.r.t. P using the strategy s_i. (i.e., $S_{s_j} \subseteq S_{s_i}$).*

The following proposition gives an exhaustive study of productivity results for the different query answering strategies studied in this paper.

Proposition 1. *Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a knowledge base with $P = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$ and $q(\vec{x})$ a query. The comparison of the query answering strategies $\{\forall, \exists, maj, \forall_c, maj_c, \exists_c\}$, with respect to productivity, is given depicted in Figure 1.*

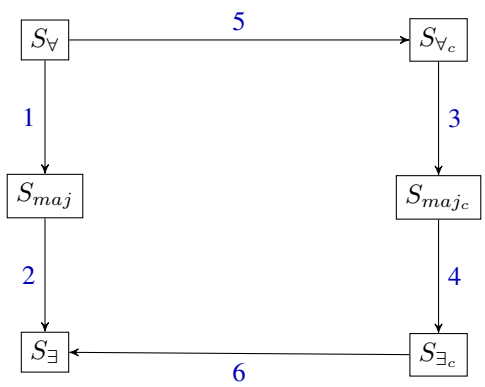


Figure 1: $X \longrightarrow Y$ means that the set of answers obtained from the strategy X is also obtained from the strategy Y

Proof. We give the proof for each relation :

- (1) The proof is immediate since $\forall s \in S_\forall$, s is such that, $\forall i \in \{1, \cdots, m\}$, $s \in S_i$, thus $\frac{|i : \vec{s} \in S_i|}{m} = 1$, $s \in S_{maj}$. (Every answer given by all ABoxes, is an answer given by the majority of ABoxes).
- (2) The proof is immediate, $S_{maj} \subseteq S_\exists$. (Every answer given by the majority of ABoxes is given by one ABox). Note that by transitivity we get $S_\forall \subseteq S_\exists$.
- (3) and (4) the proofs are analogous to (1) and (2) applied to the set of answers of maximal cardinality. By transitivity, we get that $S_{\forall_c} \subseteq S_{\exists_c}$.
- (5) The proof is immediate, $S_\forall \subseteq S_{\forall_c}$, $\forall s \in S_\forall$ by definition of Sc_P, $s \in S_i$, with $S_i \in Sc_P$ thus $s \in S_{\forall_c}$.
- (6) Since $S_\exists = S_1 \cup \ldots \cup S_m$, by definition of Sc_P, we have $S_{\exists_c} \subseteq S_\exists$.

For the converse productivity relations, we provide the following counter-examples :

Example 7. *Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a KB such that $P = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}$, with $\mathcal{A}_1 = \{A(a)\}$, $\mathcal{A}_2 = \{A(a)\}$ and $\mathcal{A}_3 = \{A(b)\}$ and let $q(x) = \exists\, x\, A(x)$ be a query.*
The sets of answers to $q(x)$ from the ABoxes are S_1 = {a}, S_2 = {a} and S_3 = {b}. Thus we have $S_\forall = \emptyset$, $S_{maj} = \{a\}$, $S_\exists = \{a, b\}$. We have (1) $S_\forall \subseteq S_{maj}$, (2) $S_{maj} \subseteq S_\exists$, and $S_\forall \subseteq S_\exists$. However, $S_{maj} \nsubseteq S_\forall$, $S_\exists \nsubseteq S_{maj}$ and $S_\exists \nsubseteq S_\forall$. Since the sets of answers have same cardinality we also have (3) $S_{\forall_c} \subseteq S_{maj_c}$, (4) $S_{maj_c} \subseteq S_{\exists_c}$ and $S_{\forall_c} \subseteq S_{\exists_c}$. However, $S_{maj_c} \nsubseteq S_{\forall_c}$, $S_{\exists_c} \nsubseteq S_{maj_c}$ and $S_{\exists_c} \nsubseteq S_{\forall_c}$.

Example 8. *Let $\mathcal{K}_P = \langle \mathcal{T}, P \rangle$ be a KB such that $P = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}$, with $\mathcal{A}_1 = \{A(a), A(c), A(d)\}$, $\mathcal{A}_2 = \{A(a), A(c), A(b)\}$ and $\mathcal{A}_3 = \{A(c), A(e)\}$ and let $q(x) = \exists\, x\, A(x)$ be a query.*
The sets of answers to $q(x)$ from the ABoxes are S_1 = { a, c, d }, S_2 = { a, c, b } and S_3 = { c, e }. Thus we have $S_\forall = \{c\}$, $S_{maj} = \{a, c\}$, $S_\exists = \{a, b, c, d, e\}$. The sets of answers to $q(x)$ of maximal cardinality are $Sc_P = \{S_1, S_2\}$, thus $S_{\forall_c} = \{a, c\}$, $S_{\exists_c} = \{a, c, d, b\}$, $S_{maj_c} = \{a, c\}$. We have (5) $S_\forall \subseteq S_{\forall_c}$ and (6) $S_{\exists_c} \subseteq S_\exists$ however $S_{\forall_c} \nsubseteq S_\forall$ and $S_\exists \nsubseteq S_{\exists_c}$.

□

Related Works

There are several lines of research related to our work. The first one is the knowledge bases merging which is a problem largely studied within the propositional logic setting. (Everaere, Konieczny, and Marquis 2010; Konieczny and Pino 2002; Lin and Mendelzon 1999; Revesz 1997; Baral, Kraus, and Minker 1991; Bloch et al. 2001) and recently in the setting of description logics where there are few works on merging (Benferhat, Bouraoui, and Loukil 2013; Benferhat et al. 2014; Moguillansky and Falappa 2007; Wang et al. 2012). These works aim to merge knowledge bases in order to get a global view which requires to know the whole content of the knowledge base before performing merging. Merging pieces of information need to use some fusion operators that permit to combine them while respecting different constraints between sources. Our paper differs from these works in two aspects : in the sense that it studies the problem of query answering without merging datasources neither knowing their contents.

The second one is concerned with the framework of databases querying sources that present some limitations has been investigated using views (Halevy 2001; Pottinger and Alon 2001; Pottinger and Halevy 2000), however these approaches mainly concern query optimization or rewriting and query languages.

The third research line related to our work is inconsistency-tolerant query answering which is handled in the framework of the database approaches (Marcelo,

Leopoldo, and Chomicki 1999; Bertossi 2011; Chomicki 2006; Decker 2010; ten Cate, Halpert, and Kolaitis 2016), propositional logic approaches(Baral, Kraus, Minker and Subrahmanian 1992; Benferhat et al. 1993; Nebel 1994; Salem, Didier, and Henri 1997) or the description logic (Lembo et al. 2010) and recently in lightweight description logic DL-Lite (Lembo et al. 2015; Bienvenu 2012; Bienvenu and Rosati 2013; Benferhat et al. 2016; Baget, et al. 2016). These works aim to solve the problem of inconsistency by mainly computing a set of consistent subsets of assertional facts called repairs, that recover consistency with respect to the ontology, so using them to carry out query answering.

the fourth research line related to our work is querying Web Data Bases without Accessing to Data (Boughammoura et al. 2017; Boughammoura, Omri and Hlaoua 2012; Boughammoura, Hlaoua and Omri 2015). These works aim to identify Web page templates and the tag structures of a document in order to extract structured data from hidden web sources as the results returned in response to a user query.

Our work differs form that, since it has been assumed that each ABox in our knowledge base is consistent w.r.t. the TBox and the responses to queries as well. However, the fact that all the datasets are consistent with the TBox, does not mean that the union of answers to a query, which is implicitly can be seen as a set of fact, will be consistent with the TBox. Let us consider the following example.

Suppose we introduce the following axioms to the TBox of the running example (Example 2), $Person \sqsubseteq Customer, TravelAgency \sqsubseteq Customer, Person \sqsubseteq \neg TravelAgency$ and consider the query $q(x) = \exists x(book(x, y) \wedge Customer(x)$. Assume that $S_\forall = \{a\}$. Since each ABox is consistent w.r.t. the TBox, it might be the case that $Person(a) \in \mathcal{A}_1$ and $Person(a) \in \mathcal{A}_2$ but $TravelAgency(a) \in A_3$, therefore the answers profile is inconsistent. In order to overcome this problem, we plan to do repair when the union of the answers to a query (for instant the universal inference) is inconsistent with the knowledge base.

Conclusion

In this paper, we investigated the query answering problem when we do not know the content of the ABox (Data) in terms of inference strategies. We proposed basic inferences from the sets of answers provided by the ABoxes. We considered the cardinality of the sets of answers in defining new cardinality-based inferences. We also compared these inferences in terms of productivity.

In a future work we plan to study the logical properties of the proposed inferences, according to the Kraus Lehman and Magidor's framework (Kraus, Lehmann, and Magidor 1990). Besides, these inference strategies can be compared according to their computational complexity which is another future issue.

Finally, as mentioned in the previous Section, another direction of investigation is how to perform query answering of a knowledge base when the multiset of ABoxes is inconsistent.

References

Alessandro, A.; Diego, C.; Roman, K.; and Michael, Z. 2009. The dllite family and relations. *J Artif Int Res* 36(1):1–69.

Baader, F.; Calvanese, D.; McGuinness, D.; Nardi, D.; and F. Patel-Schneider, P. eds. 2010. *The Description Logic Handbook Theory Implementation and Applications*. Cambridge University Press.

Baget, J.F.; Benferhat, S.; Bouraoui, Z.; Croitoru, M.; Mugnier, M.L.; Papini, O.; Rocher, S.; and Tabia, K. 2016. A general modifier based framework for inconsistency tolerant query answering. In *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning KR*.

Baral, C.; Kraus, S.; Minker, J.; and Subrahmanian, S.V. 1992. Combining knowledge bases consisting of first order theories. *Computational Intelligence* 8:45–71.

Benferhat, S.; Cayrol, C.; Dubois, D.; Lang, J.; and Prade, H. 1993. Inconsistency management and prioritized syntax based entailment. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, 640–647.

Benferhat, S.; Bouraoui, Z.; Lagrue, S.; and Rossit, J. 2014. Min based assertional merging approach for prioritized dl lite knowledge bases. 8–21.

Benferhat, S.; Bouraoui, Z.; Croitoru, M.; Papini, O.; and Tabia, K. 2016. Non objection inference for inconsistency tolerant query answering. In *IJCAI*.

Benferhat, S.; Bouraoui, Z.; Chau, M.T.; Lagrue, S.; and Rossit, J. 2017. A polynomial algorithm for merging lightweight ontologies in possibility theory under incommensurability assumption. In *International Conference on Agents and Artificial Intelligence ICAART 2017*, 415–422.

Benferhat, S.; Bouraoui, Z.; and Loukil, Z. 2013. Min based fusion of possibilistic dl-lite knowledge bases. In *Web Intelligence*, 23–28.

Bertossi, L.E. 2011. *Database Repairing and Consistent Query Answering*. Synthesis Lectures on Data Management.

Bienvenu, M.; and Rosati, R. 2013. Tractable approximations of consistent query answering for robust ontology based data access. In *IJCAI 2013 Proceedings of the 23rd International Joint Conference on Artificial Intelligence*.

Bienvenu, M. 2012. On the complexity of consistent query answering in the presence of simple ontologies. In *Proceedings of the Twenty-Sixth Conference on Artificial Intelligence*.

Bloch, E.I.; Hunter, A.; Ayoun, A.; Benferhat, S.; Besnard, P.; Cholvy, L.; Cooke, R.; Dubois, D.; and Fargier, H. 2001. Fusion general concepts and characteristics. *International Journal of Intelligent Systems* 16:1107–1134.

C, Baral.; S, Kraus.; and J, Minker. 1991. Combining multiple knowledge bases. *IEEE Trans on Knowl and Data Eng* 3(2):208–220.

Boughammoura, R.; and Omri, M.N. 2017. Querying Deep Web Data Bases without Accessing to Data. *International Conference on Natural Computation Fuzzy Systems and Knowledge Discovery*

Boughammoura, R.; Omri, M.N.; and Hlaoua, L. 2012. Information retrieval from deep web based on visual query interpretation. *International Journal of Information Retrieval Research*, IJIRR 04, 45–59

Boughammoura, R.; Hlaoua, L.; and Omri, M.N. 2015. G-Form A Collaborative Design Approach to Regard Deep Web Form as Galaxy of Concepts. *International Conference on Cooperative Design Visualization and Engineering*, CDVE, 20–23

Chomicki, J. 2006. Consistent query answering five easy pieces. In *Proceedings of the 11th International Conference on Database Theory*, ICDT 07, 1–17. Berlin and Heidelberg: Springer Verlag.

Decker, H. 2010. Basic causes for the inconsistency tolerance of query answering and integrity checking. In *2010 Workshops on Database and Expert Systems Applications*, 318–322.

Halevy, Y. A. 2001. Answering queries using views a survey. *The VLDB Journal* 10(4):270–294.

Lin, J.; and Mendelzon, A. 1999. *Knowledge Base Merging by Majority*. Dordrecht: Springer Netherlands. 195–218.

Kraus, S.; Lehmann, D.J.; and Magidor, M. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artif. Intell.* 44(1-2):167–207.

Lembo, D.; Lenzerini, M.; Rosati, R.; Ruzzi, M.; and Savo, D.F. 2010. Inconsistency tolerant semantics for description logics. In *Web Reasoning and Rule Systems Fourth International Conference RR 2010*, 103–117.

Lembo, D.; Lenzerini, M.; Rosati, R.; Ruzzi, M.; and Savo, D.F. 2015. Inconsistency tolerant query answering in ontology based data access. *J Web Sem* 33:3–29.

Marcelo, A.; Leopoldo, B.; and Chomicki, J. 1999. Consistent query answers in inconsistent databases. In *Proceedings of the Eighteenth ACM SIGMOD SIGACT SIGART Symposium on Principles of Database Systems*, PODS 99, 68–79. New York and NY and USA: ACM.

Maurizio, L. 2011. Ontology based data management. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM 11, 5–6. New York NY USA: ACM.

Moguillansky, M.O., and Falappa, M.A. 2007. A non monotonic description logics model for merging terminologies. *Inteligencia Artificial Revista Iberoamericana de Inteligencia Artificial* 11(35):77–88.

Muro, M.R.; Kontchakov, R.; and Zakharyaschev, M. 2013. Ontology based data access ontop of databases. 558–573.

Nebel, B. 1994. Base revision operations and schemes semantics representation and complexity. In *Proc 11th European Conference on Artificial Intelligence*, 341–345. John Wiley Sons.

Everaere, P.; Konieczny, S.; and Marquis, P. 2010. Disjunctive Merging Quota and Gmin Merging Operators. *Artificial Intelligence*. To appear.

Poggi, A.; Lembo, D.; Calvanese, D.; Giacomo, G.D.; Lenzerini, M.; and Rosati, R. 2008. Linking data to ontologies. *J Data Semantics* 10:133–173.

Pottinger, R.; and Alon, H. 2001. Minicon a scalable algorithm for answering queries using views. *The VLDB Journal* 10(2-3):182–198.

Pottinger, R.; and Halevy, A. 2000. A scalable algorithm for answering queries using views. In *Proceedings of the 26th International Conference on Very Large Data Bases*, VLDB00, 484–495. San Francisco and CA and USA: Morgan Kaufmann Publishers Inc.

Reiter, R. 1987. A theory of diagnosis from first principles. *Artif Intell* 32(1):57–95.

Revesz, Z.P. 1997. On the semantics of arbitration. *International Journal of Algebra and Computation* 07(02):133–160.

Salem, B.; Didier, D.; and Henri, P. 1997. Some syntactic approaches to the handling of inconsistent knowledge bases a comparative study part 1 the flat case. *Studia Logica* 58(1):17–45.

Konieczny, S.; and Pino, P.R. 2002. On the frontier between arbitration and majority. In *Proceedings of the Eights International Conference on Principles of Knowledge Representation and Reasoning*, KR02, 109–120. San Francisco CA USA: Morgan Kaufmann Publishers Inc.

ten Cate, B.; Halpert, R.; and Kolaitis, P. 2016. Practical query answering in data exchange under inconsistency tolerant semantics. In *EDBT*, 233–244.

Wang, Z.; Wang, K.; Jin, Y.; and Qi, G. 2012. Ontomerge a system for merging dl-lite ontologies.