A Fuzzy Datalog Deductive Database System

Pascual Julián-Iranzo, Fernando Sáenz-Pérez

Abstract—This paper describes a proposal for a deductive database system with fuzzy Datalog as its query language. Concepts supporting the fuzzy logic programming system Bousi-Prolog are tailored to the needs of the deductive database system DES. We develop a version of fuzzy Datalog where programs and queries are compiled to the DES core Datalog language. Weak unification and weak SLD resolution are adapted for this setting, and extended to allow rules with truth degree annotations. We provide a public implementation in Prolog which is open-source, multiplatform, portable, and in-memory, featuring a graphical user interface. A distinctive feature of this system is that, unlike others, we have formally demonstrated that our implementation techniques fit the proposed operational semantics. We also study the efficiency of these implementation techniques through a series of detailed experiments. Moreover, a database example for a recommender system is used to illustrate some of the features of the system and its usefulness.

Index Terms—Deductive Database, Fuzzy Logic Programming, Fuzzy Prolog, Weak Unification, Bousi-Prolog, Datalog Educational System.

I. INTRODUCTION

Fuzzy Logic Programming integrates concepts coming from fuzzy logic [1] into pure logic programming [2] in order to deal with imprecise information, uncertainty and/or vagueness by using declarative techniques. In the last decade there has been a renewed interest in this amalgamation as revealed in the works [3], [4]. There is not a common method for introducing fuzzy concepts into logic programming. In particular, Bousi-Prolog [5], [6], an extension of the Prolog language, follows the approach of [7], where the syntactic unification mechanism of classical SLD resolution is replaced by a weak unification algorithm, based on proximity/similarity relations. This algorithm provides a weak most general unifier as well as a numerical value, called the approximation degree. Intuitively, the approximation degree represents the truth degree associated with the (query) computed instance. Programs written in this language consist, in essence, of a set of ordinary (Prolog) clauses jointly with a set of “proximity equations” which play an important role during the unification process.

Datalog [8] is a query language for deductive databases that can be seen as a syntactic subset of Prolog. Pure Datalog is a truly declarative language because the order neither of rules nor goals in the program do affect operational semantics (in particular, non-logic constructors are disallowed, as the cut), but it is not Turing-complete as it is meant as a database query language.

Fuzzy Datalog [9] is an extension of Datalog-like languages using lower bounds of uncertainty degrees in facts and rules. Akin proposals as [10], [11] explicitly include the computation of the rule degree, as well as an additional argument to represent this degree. However, and similar to Bousi-Prolog, we are interested in removing this burden from the user with an automatic rule transformation that slides both the degree argument and the explicit call to fuzzy connective computations in user rules. In addition, we provide support for several proximity/similarity relations such as in Bousi-Prolog.

Thus, in this paper, we are interested in the implementation of a deductive fuzzy database with such features by extending the Datalog Educational System (DES) [12] into a system that we call FuzzyDES from here on. The DES system is a deductive database targeted at teaching databases and their query languages. We focus on Datalog to be extended with fuzzy relations (i.e., relations where each data tuple is associated to an approximation degree), and weak unification and resolution. By contrast to a fuzzy Prolog system, answers are (multi)-sets instead of just single answers retrieved via backtracking. Whereas queries solved under SLD resolution can easily develop non-termination, queries under tabled SLD resolution [13] are ensured to terminate for user predicates, which is a natural requirement in the database arena (cf. SQL). Due to the interactive nature of DES, we keep this feature in the fuzzy setting to let users play and interact with the system by providing commands to assert new rules and facts on the fly.

For the implementation, we transfer the techniques developed for the Bousi-Prolog system into the DES system. In general, following the approach of the high level implementation of Bousi-Prolog, a FuzzyDES program is compiled into a set of core Datalog rules, which in turn are interpreted by a deductive engine implemented in Prolog. The translation includes calls to a collection of auxiliary predicates able to reproduce the Weak SLD resolution with graded rules (i.e., annotated rules with weights acting as truth degrees). Since Datalog differ from Prolog (e.g., non-compound terms and ground answers), we adapt and specialize the Weak SLD resolution (WSLD) procedure found in the Bousi-Prolog system to DES. A distinctive feature of the built system is that, unlike others, we have formally demonstrated that our techniques for implementing the WSLD resolution operational semantics are correct, i.e., they produce the expected answers and only them. We also study the efficiency of these implementation techniques through a series of detailed experiments.

To the best of our knowledge, there is no a publicly-available implementation of a fuzzy Datalog system as the
one we propose in this work. This paper describes our approach to develop such a system. Our motivation lies in that, although SQL (as a query language) dominates the database panorama, over the years SQL has become quite complex and some statements have started to cause problems for end users and even for professionals. Also, one of the main advantages of deductive databases is the ability to specify recursive rules without the limitations present in SQL systems (linearity and termination control). This, jointly with the addition of fuzzy features, can facilitate the flexible use of databases and knowledge bases. Additionally, in the last years we have witnessed an increased interest in recursive Datalog queries in a variety of application domains such as data integration, information extraction, networking, program analysis, security, and even cloud computing (cf. Section III). Finally, it is important to emphasize that our system has been specially designed to facilitate flexibility for fuzzy querying and answering of questions, encapsulating fuzzy management. This is a definitive feature of our system that do not have others based on more traditional techniques, in which the user has to set certain parameters (as truth degrees and their manipulation) when specifying their queries and rules. This is an inherited characteristic of Bousi-Proplog that allows a complete separation when specifying logic, vague knowledge and control. All of these are appreciable features that facilitate the modelling of knowledge systems from a more declarative perspective.

The following section gives a practical motivation to this work by describing an application example of FuzzyDES showing its ability to model uncertain knowledge and facilitate flexible query answering.1

II. A MOTIVATING EXAMPLE

Recommender systems are an effective way to assist people by offering advice on finding suitable products and services to facilitate online decision-making [14]. Subjective information (such as how a person evaluates a product or service, and how another person trust an opinion depending on its confidence) can be specified with linguistic, fuzzy information. In order to motivate our work, as an example of a fuzzy database in DES, we apply these ideas to modelling a small recommender system for restaurants in Madrid. We consider people interested in asking questions relating the location of the restaurant with respect to its own location, the quality of the restaurant in terms of other’s opinions, and the type of food served. Thus, we can distinguish two fuzzy relations: A proximity relation near (reflexive and symmetric) for representing walking distances, and a predefined similarity relation ~ (which is, in addition, transitive) for representing quality degrees. These are defined in the next code excerpt of the database file that can be downloaded from http://des.sourceforge.net/fuzzy/recommender.dl.

\[
\begin{align*}
&\text{fuzzy_relation(near, [reflexive, symmetric])}. \\
&\text{sol near callao} = 0.6. \quad \text{sol near cruz} = 0.5. \\
&\text{callao near plaza_espàna} = 0.4. \\
&\text{fuzzy_relation(~, [reflexive, symmetric, transitive])}.
\end{align*}
\]

The first assertion fuzzy_relation defines the operator near as a proximity relation for a walking distance with the reflexive and symmetric properties. Analogously, a similarity relation ~ is explicitly defined next (in fact, its assertion can be removed since it is defined as such by default). Proximity equations, as plain~good=0.5, specify the degree of semantic similarity for two different syntactic symbols. In this particular case it states that plain and good are similar with approximation degree 0.5.

A fuzzy relation confidence/1 defines the degree of reliability (in the opinion) which a given type of user deserves. For example, a local guide is assumed to be a person including serious comments in the database where normal and casual users may not, and with different confidences:

\[
\begin{align*}
&\text{confidence(local_guide) with 0.9.} \\
&\text{confidence(normal_user) with 0.5.} \\
&\text{confidence(casual_user) with 0.3.}
\end{align*}
\]

Next, several relations are defined with facts in the database: The relation restaurant/3 relates the name of a restaurant, its location, and food served. Users and their types are related in the relation user/3. Finally, user comments are represented in the relation comment/3 relating user, restaurant, and comment.

\[
\begin{align*}
&\text{restaurant(don_osco,cruz,burgerer).} \\
&\text{restaurant(roque,solrice).} \\
&\text{restaurant(tagliatella,benavente,italian).} \\
&\text{user(juan,local_guide).} \\
&\text{user(sara,normal_user).} \\
&\text{user(pepe,casual_user).} \\
&\text{comment(juan,don_osco,plain).} \\
&\text{comment(juan,rodilla,good).} \\
&\text{comment(pepe,roque,excellent).} \\
&\text{comment(sara,tagliatella,very_good).}
\end{align*}
\]

A recommendation relies on the quality of a restaurant, which is defined with a single rule that takes into account the comment a user has provided, the type of this user, and its confidence:

\[
\begin{align*}
&\text{quality(Restaurant,Quality) :-} \\
&\text{comment(User,Restaurant,Quality),} \\
&\text{user(User,Type), confidence(Type).}
\end{align*}
\]

So, in order to provide recommendations, the rule recommend relates restaurants (Restaurant) with the location of the user (Origin), serving certain food (Food) with an acknowledged quality (Quality):

\[
\begin{align*}
&\text{recommend(Origin,Food,Quality,Restaurant) :-} \\
&\text{restaurant(Restaurant,Location,Food),} \\
&\text{Location near Origin, quality(Restaurant,Quality).}
\end{align*}
\]

Contrary to Fuzzy Prolog implementations, the set-oriented approach of deductive databases makes queries to return all the answers, similar to relational databases, so that a query to this database can return several recommendations with different approximation degrees at once. For instance, the query in the following system session returns recommendations for a user located at sol:

DEDS> /consult recommender
Info: 22 rules consulted.
In this session, after switching the deductive system to the fuzzy setting, the command consult loads the database located in the given file (with default extension .dl). The query returns all possible restaurants indicating the type of food served, quality and the recommended restaurant. Each answer tuple is ordered by default with respect to descending approximation degrees. The first tuple indicates that rodilla has a support degree of 0.6, which is a value constructed by taking into account that the restaurant is located at callao, which is near sol. Also, there is a comment from the local guide juan (with a support of 0.9) stating that the restaurant is good. The next tuple in the answer refer to the restaurant don_oso, receiving a support degree of 0.5 with quality plain, because the same local guide commented on this restaurant with this quality level for a restaurant located at cruz, near sol. The last tuple receives a small support because the comment was raised by a user whose type is scored rather low.

We can interactively add information about the type of the food served with:

```prolog
FDES> /assert burguer~fast_food=0.7.
FDES> /assert snacks~fast_food=0.9.
```

Then, a similar query but looking for good fast food near sol (with no need to reload the database) would retrieve:

```prolog
FDES> recommend(sol,fast_food,good,Restaurant)
```

There are really two tuples for don_oso fulfilling the question, with the same support degree (one for plain and other for good quality). Due to the default set-oriented behavior of the system, duplicates are removed. However, duplicates can be enabled with the command /duplicates on, and the same query would return 3 tuples.²

An approximation degree threshold (λ-cut) can be stated with the command /lambda_cut Value, which prunes computations as soon as a degree greater than the threshold is computed as a result of an application of the t-norm $\Delta$, which is a binary truth function generalizing classical conjunction. In this last example, the second answer would be removed from the answer for a λ-cut of 0.55.

Facts and rules can also be interactively added (with the command /assert Rule or from a file with /reconsult File) so that the database have not to be recompiled each time it is modified. Finally, a query for the predicate recommend with variables in all its arguments would return all possible recommendations. (The reader is encouraged to try the system with different queries.)

²Nonetheless, as it will explained later in Subsection IV-F, enabling answer subsumption automatically removes output tuples with equal or lesser approximation degrees.

The Datalog Educational System (DES) [12] is a free, open-source, multiplatform, portable, in-memory, Prolog-based implementation of a deductive database system. DES 5.0.1 (des.sourceforge.net) is the current release, which includes several query languages: Datalog, SQL, Relational Algebra and Relational Calculi. Fuzzy Datalog is the main addition for this new major release. DES features tabling, types, integrity constraints, stratified negation [8], persistency, full-fledged arithmetic, ODBC connections to relational databases, novel approaches to Datalog and SQL declarative debugging [15], [16], test case generation for SQL views [17], null value support, outer join, and aggregate predicates and functions [12]. This system is used world-wide in many universities (e.g. Imperial College London –UK–, TU München –Germany–, Université Lille 1 –France–, UCLA –USA– or The University of Sydney –Australia–) for teaching deductive databases (see http://des.sourceforge.net/html/what_s_des_for_html, in its web page, for an extensive list of institutions), and used as a test-bed for research (Mozilla leaks, smart grid networks, deductive data warehouses, declarative debugging, . . . ), and featuring more than 71K downloads up to now.

### III. Datalog Educational System

A deductive database in DES is defined by a normal logic program with rules of the form $A \leftarrow Q$, where the body $Q \equiv L_1 \land \ldots \land L_n$ is a query, the head $A$ is an atom and, each $L_i$ is a literal. In this paper, we restrict ourselves to definite clauses (with positive literals). Thus, a literal can be a call to either a defined predicate or to a built-in predicate. A disjunctive rule $A \leftarrow Q_1 \lor Q_2$ is understood as the rules $A \leftarrow Q_1$ and $A \leftarrow Q_2$. Textual syntax in the system follows Prolog convention: Variables start with either upper-case or an underscore, and predicate symbols and string constants either start with down-case or are delimited by single quotes. The left implication $\leftarrow$ is written as :-, the conjunction with a comma (,), and the disjunction with a semicolon (;). We interchangeably use the terms: predicate - relation, clause - rule, and goal - query.

#### A. Syntax

A deductive database in DES is defined by a normal logic program with rules of the form $A \leftarrow Q$, where the body $Q \equiv L_1 \land \ldots \land L_n$ is a query, the head $A$ is an atom and, each $L_i$ is a literal. In this paper, we restrict ourselves to definite clauses (with positive literals). Thus, a literal can be a call to either a defined predicate or to a built-in predicate. A disjunctive rule $A \leftarrow Q_1 \lor Q_2$ is understood as the rules $A \leftarrow Q_1$ and $A \leftarrow Q_2$. Textual syntax in the system follows Prolog convention: Variables start with either upper-case or an underscore, and predicate symbols and string constants either start with down-case or are delimited by single quotes. The left implication $\leftarrow$ is written as :-, the conjunction with a comma (,), and the disjunction with a semicolon (;). We interchangeably use the terms: predicate - relation, clause - rule, and goal - query.

#### B. Safe Databases

Though the Datalog language might be understood as a syntactic subset of Prolog, it is claimed to be a true declarative language (see, e.g., [18]) in the sense that rule and goal ordering in a program does not affect declarative semantics. However, as a database language, users expect termination and finite answers, so that some restrictions to programs are imposed. The first restriction is about safe rules [8], [19], that ensure closed (ground) answers. Open answers (i.e., including unbound variables) are not allowed as they can represent infinite tuples (for example, the answer $p(X)$ represents $\forall X p(X)$, a domain which is not defined as finite in the database). A second restriction refers to built-ins because whereas user-defined predicates represent finite relations, some useful built-ins represent infinite relations. For example, a comparison operator $(\leq, >, \ldots)$ relates two arguments representing a true relation for all the possible elements fulfilling the comparison...
D. Tabling

where {X} the same value for group_by(r(X,Y), [X], M=max(Y)) including in general an aggregate. For example, the query E as a grouping criterion, and at p stratum number to predicate given iteration of a fixpoint computation, all the tuples for aggregate can be a source of non-monotonicity because in a of this technique for computing aggregates. Computing an aggregate can be a source of non-monotonicity because in a given iteration of a fixpoint computation, all the tuples for the relation on which the aggregate operates are not available in general. So, there exists the risk of deducing an incorrect outcome in a given iteration that can be neglected in a further iteration. For instance, we can assume that the tuple r(0) is in the current fixpoint iteration, so that computing the maximum of the single argument of r yields 0. But it is possible that in a next iteration, a new tuple (say r(1)) is computed, therefore neglecting the previous computed value for the aggregate. Thus, stratification is applied to ensure that all the tuples of the relation r are computed before trying to compute the aggregate.

A stratification for a program is built with the aid of a predicate dependency graph (PDG) [19], showing both the positive and negative dependencies between predicates in the program. Each node in this graph is a program predicate symbol, and there are as many nodes as such symbols in the program. Arcs come from each predicate in a rule body (antecedent) to its rule predicate (consequent). If the antecedent occurs as an aggregate argument, its arc is labelled as negative, and positive otherwise. A stratification collects predicates into numbered strata so that, given the function \( str(\Pi, p) \) which assigns a stratum number to predicate \( p \) in a database \( \Pi \), then for a positive arc \( p \leftarrow q \), \( str(\Pi, p) \leq str(\Pi, q) \), and for a negative arc \( p \leftarrow q \), \( str(\Pi, p) < str(\Pi, q) \).

In this paper, we use the aggregate \( \text{group}_by(R, Vs, E) \), where \( R \) is a relation call, \( Vs \) are the variables in \( R \) used as a grouping criterion, and \( E \) is a Boolean expression including in general an aggregate. For example, the query \( \text{group}_by(r(X,Y), [X], M=\max(Y)) \) is intended to compute the maximum value of \( Y \) for each group formed by tuples with the same value for \( X \). Thus, if the meaning of \( r \) is defined by \{ r(a,0), r(a,1), r(b,3) \}, then the previous query returns \{ \text{answer}(a,1), \text{answer}(b,3) \}. 

C. Stratification and Aggregates

Deductive system implementations require some form of fixpoint computation to deduce the meaning of programs and queries. Keeping this computation monotonic is achieved by the well-known technique called stratification [8], which is as a syntactic restriction that rejects programs which combine both negation and recursion in a computation path. Though in our work we do not deal with negation, we still take advantage of this technique for computing aggregates. Computing an aggregate can be a source of non-monotonicity because in a given iteration of a fixpoint computation, all the tuples for the relation on which the aggregate operates are not available in general. So, there exists the risk of deducing an incorrect outcome in a given iteration that can be neglected in a further iteration. For instance, we can assume that the tuple r(0) is in the current fixpoint iteration, so that computing the maximum of the single argument of r yields 0. But it is possible that in a next iteration, a new tuple (say r(1)) is computed, therefore neglecting the previous computed value for the aggregate. Thus, stratification is applied to ensure that all the tuples of the relation r are computed before trying to compute the aggregate.

A stratification for a program is built with the aid of a predicate dependency graph (PDG) [19], showing both the positive and negative dependencies between predicates in the program. Each node in this graph is a program predicate symbol, and there are as many nodes as such symbols in the program. Arcs come from each predicate in a rule body (antecedent) to its rule predicate (consequent). If the antecedent occurs as an aggregate argument, its arc is labelled as negative, and positive otherwise. A stratification collects predicates into numbered strata so that, given the function \( str(\Pi, p) \) which assigns a stratum number to predicate \( p \) in a database \( \Pi \), then for a positive arc \( p \leftarrow q \), \( str(\Pi, p) \leq str(\Pi, q) \), and for a negative arc \( p \leftarrow q \), \( str(\Pi, p) < str(\Pi, q) \).

In this paper, we use the aggregate \( \text{group}_by(R, Vs, E) \), where \( R \) is a relation call, \( Vs \) are the variables in \( R \) used as a grouping criterion, and \( E \) is a Boolean expression including in general an aggregate. For example, the query \( \text{group}_by(r(X,Y), [X], M=\max(Y)) \) is intended to compute the maximum value of \( Y \) for each group formed by tuples with the same value for \( X \). Thus, if the meaning of \( r \) is defined by \{ r(a,0), r(a,1), r(b,3) \}, then the previous query returns \{ \text{answer}(a,1), \text{answer}(b,3) \}.
Here we give our slightly different version of it, which is algorithm coincides with the one defined by M. Sessa [7]. When we work with similarity relations, this Prolog variables are consider similar with approximation degree distinct different syntactic symbols represent distinct infor-

A. Fuzzy Relations and Weak Unification

A binary fuzzy relation on a set U is a fuzzy subset on $U \times U$ (that is, a mapping $U \times U \rightarrow [0,1]$). There are some important properties that may fuzzy relations may have: (1) (Reflexive) $R(x,x) = 1$ for any $x \in U$; (2) (Symmetric) $R(x,y) = R(y,x)$ for any $x,y \in U$; (3) (Transitive) $R(x,z) \geq R(x,y) \land R(y,z)$ for any $x,y,z \in U$; where the operator $\triangle$ is an arbitrary t-norm. The notion of transitivity above is $\triangle$-transitive, and if the operator $\triangle$ is the minimum of two elements (Gödel t-norm), we speak of min-transitive.

A proximity relation is a binary fuzzy relation which is reflexive and symmetric. A proximity relation which in addition fulfills the transitive property is called a similarity relation.

We are primarily interested in similarity relations on the alphabet of a first order language. The reason is to break the usual constraint in programming languages by which different syntactic symbols represent distinct information. So, in this context, two different constant symbols or predicate/(function) symbols, with the same arity, can be treated as equal up to a certain degree. Syntactically equal variables are consider similar with approximation degree 1; otherwise, their approximation degree is 0.

For a Datalog language, where a term can only be either a constant or a variable (i.e., no function symbols are allowed), the similarity relation $R$ on the symbols of the alphabet can be extended to atomic formulas as follows: Let $p$ and $q$ be two $n$-ary predicate symbols and let $t_1, \ldots, t_n, s_1, \ldots, s_n$ be either constants or variables, then,

$$R(p(t_1, \ldots, t_n), q(s_1, \ldots, s_n)) = R(p,q)\triangle(\Delta_{i=1}^{n} R(t_i, s_i)).$$

FuzzyDES inherits the weak unification algorithm of Bousi~Prolog. When we work with similarity relations, this algorithm coincides with the one defined by M. Sessa [7]. Here we give our slightly different version of it, which is thresholded by a cut value $\lambda \in [0,1]$, also known as $\lambda$-cut.

In the context of a similarity relation it is possible to define a weak notion of most general unifier. Let $R$ be a similarity relation, $\lambda$ be a cut value and $E_1$ and $E_2$ be two expressions. The substitution $\theta$ is a weak unifier of level $\lambda$ for $E_1$ and $E_2$ with respect to $R$ (or $\lambda$-unifier) if its unification degree $R(E_1, E_2\theta) \geq \lambda$. A substitution $\theta$ is a weak most general unifier (wmgu) of level $\lambda$ (or $\lambda$-wmgu), w.r.t. $R$, for $E_1$ and $E_2$, denoted by $wmgu_R(E_1, E_2)$, if: (1) $\theta$ is a lambda-unifier of $E_1$ and $E_2$; and (2) for any lambda-unifier $\sigma$ of $E_1$ and $E_2$, the substitution $\theta$ is more general than the substitution $\sigma$ with level $\lambda$; that is, there exists a substitution $\delta$ such that, for any variable $x$ in $Dom(\sigma) \cup Dom(\theta)$, $R(x, x\delta) \geq \lambda$.

Weak most general unifiers are computed by means of a weak unification algorithm which is formalized as a transition system supported by a similarity-based unification relation $\Rightarrow$. For a similarity relation $R$ and a cut value $\lambda$, the unification of the expressions $E_1$ and $E_2$ is obtained by a state transformation sequence starting from an initial state $\langle G_0 \equiv \{E_1 \approx E_2\}, id, \alpha_0 \rangle$, where $id$ is the identity substitution and $\alpha_0 = 1$: $(G_n, id, \alpha_n) \Rightarrow (G_{n+1}, \theta_n, \alpha_n)$. When $G_n = \emptyset$ and $\alpha_n \geq \lambda$, the expressions $E_1$ and $E_2$ are unifiable by similarity with $\alpha_n$ and unification degree $\alpha_n$. Otherwise, if $G_n$ = Fail, $E_1$ and $E_2$ fail to unify.

The similarity-based unification relation ($\Rightarrow$) is defined as in the classical unification algorithm, except for the term decomposition and failure rules (see later in Subsection IV-C and in [6] for a more formal and extensive discussion of this topic). Note that, unlike [7], the resulting approximation degree is limited by a cut value $\lambda \geq 0$. Nevertheless, the weak unification theorem proved in [7, pag. 412] is valid in our framework.

B. Implementing Fuzzy Relations

As in Bousi~Prolog, a fuzzy relation $R$ can be specified by stating, first, a set of what we called relationship equations and, second, its properties. A relationship equation is a program declaration with textual form $x/a \equiv y/b = a$, which represents entries $R(x,y) = \alpha$ of the fuzzy binary relation $R$ (textually represented by an infix operator $\equiv$), and its intuitive reading is that two $n$-ary symbols, $x$ (x) and $y$ (y), are related with a certain degree $\alpha$ (a), where the arity specification for constants ($/0$) is omitted in the equation. Note that when the relation partially specified by the relationship equations is a proximity (or a similarity) we will often speak of proximity equations. The relationship equations specifying the standard similarity relation are represented internally by atoms $\forall (x,y,\alpha)$ meaning that two symbols $X$ and $Y$ are related with the approximation degree $\alpha$.

The properties attached to a relation $R$ represent its intensional description, so that users are not expected to include all the relationship equations fulfilling such properties. For instance, obtaining the (additional) entries in a similarity relation $R$ (which are needed for the compilation of e-clauses –see Subsection IV-E–) is performed by automatically computing its $\triangle$-closure [23, 24]. This closure can be understood as the shortest paths in a weighted directed graph (cf. Floyd-Warshall algorithm), where the length of a path $x_1 \alpha_1 \rightarrow x_2 \alpha_2 \rightarrow \cdots \alpha_{n-1} x_n$ is $\Delta_{i=1}^{n-1} \alpha_i$. By taking advantage of the DES deductive engine, this closure can be neatly specified with Datalog rules for each of the properties with no need to resorting to a specific implementation of a $\triangle$-closure algorithm.
In the concrete implementation of FuzzyDES, a default similarity relation \( R \) is denoted by the identifier \(-\) and it is represented internally by a set of atoms \(- (X,Y,D)\) relating two symbols \( X \) and \( Y \) with the approximation degree \( D \). The similarity relation \(-\) is intensionally defined by its properties with the following Datalog rules, where underscored variables are non-relevant variables for the outcome:\(^4\)

\[
\begin{align*}
- (X,1.0) &\leftarrow - (X,_,_D) \mid - (_,Y,_,D) . \\
- (X,Y,D) &\leftarrow - (Y,X,D) . \\
- (X,Y,D) &\leftarrow - (X,Z,D_1), - (Z,Y,D_2), \\
&\quad \text{\( t\_\text{norm}(\sim, [D_1,D_2], D) . \)} \\
- (X,Y,D) &\leftarrow \text{\( \text{\text{group\_by}}(- (X,Y,D_1), [X,Y], D=\text{\text{max}}(D_1)). \)}
\end{align*}
\]

where the call to the predicate \( t\_\text{norm}/3\) represents the application of the t-norm \( \triangle \) associated to a relation \( R \) (e.g., the relation \(-\) has associated the minimum t-norm by default), provided in its first argument, to the list of approximation degrees given in the second argument. The result of this operation is returned in the third argument of \( t\_\text{norm}/3\). The call to the aggregate metapredicate \( \text{\text{group\_by}}/3\) groups tuples in \(-\) by the criteria \( X,Y \), and applies the aggregate expression \( D=\text{\text{max}}(D_1) \) over the partitioned relation (cf. extended relational algebra operation [25]). Note that \( \sim/2 \) in DES does not just represent unification, but performs expression evaluation of both arguments followed by their unification (compound terms are not allowed as data). In plain words, this operation selects among the different \(- (X,Y,D_1)\) with the same arguments \( X \) and \( Y \), the one which has maximum \( D_1 \).

Solving the aggregate in the metapredicate \( \text{\text{group\_by}}/3\) requires that the relation \(-\), in its first argument, must be placed in a lower stratum (similarly to what happens with the negation) than the one of the predicate \( (\sim)\) containing the rule with the metapredicate. For a relation with no transitive property, its closure coincides with the classical closure, and the last rule simply becomes: \( - (X,Y,D) \leftarrow - (X,Y,D) \), so that both \(-\) and \( \sim\) are in the same stratum. Finally, note that only some built-ins as this grouping can contain compound terms and goals as arguments.

The system includes several commands for stating the properties and t-norm of a given relation. In particular, the command \( /\text{\text{fuzzy\_relation Relation ListOfProperties}}\) sets the relation name with its properties given as a list (including reflexive, symmetric and transitive). The command \( /t\_\text{norm Relation TNorm} \) sets the t-norm (goedel, lukasiewicz, product, hamacher, nilpotent, where min is synonymous for goedel, and luka for lukasiewicz) for the given relation.

### C. Implementing Weak Unification

Weak unification applies both to terms and predicates. The specific weak unification algorithm is implemented following closely Martelli and Montanari’s unification algorithm for syntactic unification [26]. The occurs check is not needed in the deductive setting because compound terms are not allowed (as \( X=f(X) \) in Prolog, which would represent an infinite structure).

\[
\begin{align*}
\text{\text{weak\_unify(Atomic1, Atomic2, Lambda, Degree)} :}&\rightarrow \\
&\quad \text{\( \text{\text{atomic}}(\text{\text{Atomic1}}), \text{\text{atomic}}(\text{\text{Atomic2}}), 1, \)} \\
&\quad \text{\( \text{\text{unification\_degree}}(\text{\text{Atomic1}}, \text{\text{Atomic2}}, \text{\text{Degree}}), \)} \\
&\quad \text{\( \text{\text{Degree}} \gg \text{\text{Lambda}}. \)} \\
\text{\text{weak\_unify(Term, Variable, \_Lambda, 1.0)} :}&\rightarrow \\
&\quad \text{\( \text{\text{nonvar}}(\text{\text{Term}}), \text{\text{var}}(\text{\text{Variable}}), 1, \text{\text{Variable}} = \text{\text{Term}}. \)} \\
\text{\text{weak\_unify(Variable, Term, \_Lambda, 1.0)} :}&\rightarrow \\
&\quad \text{\( \text{\text{var}}(\text{\text{Variable}}), \text{\text{Variable}} = \text{\text{Term}}. \)}
\end{align*}
\]

The first clause returns the top approximation degree for unifying a constant with itself if the relation is reflexive.\(^5\) Otherwise it either returns the degree between them (if exists) or fails. The final cut removes alternatives and provides a way to select the representative of the wmu class. Hence, this implementation provides a weak most general unifier as well as a numerical value, called the unification degree in [7]. Intuitively, the unification degree will represent the truth degree associated with the (query) computed instance.

FuzzyDES implements a weak unification operator, denoted by \( \sim \), which is the fuzzy counterpart of the syntactical unification operator of standard Prolog. It can be used, first, to unify two terms as in the goal Term1\( \sim\)Term2, and, second, to construct expressions. So, the expression Term1\( \sim\)Term2 returns the unification degree when evaluated, and can be used wherever an expression is expected in a goal, as in 1\( \sim\)0.5 (if \( a\sim b=0.4 \), then the goal succeeds because \( 1 - 0.4 > 0.5 \)).

Solving a goal \( \text{\text{Expr1 Op Expr2}} \) in DES, where \( \text{\text{Op}} \) is a comparison operator (\( =, \neq, >, >\neq, =\neq \) and each \( \text{\text{Expr}} \) is an expression, amounts to evaluate both expressions and comparing them with respect to the operator (with the exception of \( = \), which performs classical unification on evaluated terms).

### D. FuzzyDES Programs and Weak SLD Resolution

FuzzyDES defines a program II as a fuzzy theory, that is, as a mapping applying a finite set of formulas, namely rules, into the elements (truth values) of the lattice \( [0,1] \). Informally, a FuzzyDES program can be seen as a set of pairs \( \langle R;\alpha \rangle \), where \( R \) is a rule and \( \alpha = \Pi(R) \) is a truth degree expressing the confidence which the system user has in the truth of the rule \( R \). We call such a pair a graded rule. For rules, FuzzyDES follows the same syntactical conventions as described in Subsection III-A, and textual syntax of a graded rule \( \langle R;\alpha \rangle \) is the textual syntax for \( R \) followed by \( \gg \alpha \).

\(^4\)This clause is introduced for efficiency reasons, by omitting the search for entries \( - (\text{\text{Atomic1}}, \text{\text{Atomic2}}, 1.0) \) in the \( \Delta \)-closure. These extensional entries could be omitted as well.

\(^5\)This clause is introduced for efficiency reasons, by omitting the search for entries \( - (\text{\text{Atomic1}}, \text{\text{Atomic2}}, 1.0) \) in the \( \Delta \)-closure. These extensional entries could be omitted as well.
The next definition, which formalizes FuzzyDES operational semantics, enhances the definition in Subsection 5.2 of [6] to deal with graded rules.

**Definition 4.1:** Let II be a FuzzyDES program, $\mathcal{R}$ a similarity relation on the first order alphabet induced by II, $\Delta$ the fixed t-norm associated to $\mathcal{R}$, and $\lambda$ a $\lambda$-cut. We define $\text{Weak SLD (WSLD) resolution}$ as a transition system $(E, \Rightarrow_{\text{WSLD}})$ where $E$ is a set of triples $\langle G, \theta, \alpha \rangle$ (goal, substitution, approximation degree), that we call the state of a computation, and whose transition relation $\Rightarrow_{\text{WSLD}} \subseteq (E \times E)$ is the smallest relation which satisfies:

$$\langle \langle A' \land Q' \rangle, \theta, \alpha \rangle \Rightarrow_{\text{WSLD}} \langle \langle Q \land Q' \rangle, \sigma, \theta \sigma, \delta \Delta \beta \alpha \rangle$$

if $R \equiv \langle A \leftarrow Q \rangle; \delta \ll II, \Delta \geq \sigma, \text{wmu}_E^\lambda(A, A') \neq \text{fail}, \beta = \mathcal{R}(\sigma \theta \sigma, \alpha) \geq \lambda$, and $(\Delta \beta \delta \alpha \sigma \eta) \geq \lambda$.

Where, $Q$ and $Q'$ are conjunctions of atoms and the notation $\ll$ represents that $R$ is a standardized apart rule in II.

A WSLD derivation for $\Pi \cup \{G_0\}$ is a sequence of WSLD resolution steps $\langle G_0, \text{id}, 1 \rangle \Rightarrow_{\text{WSLD}} \langle G_1, \text{id}, 0 \alpha \rangle \Rightarrow_{\text{WSLD}} \ldots \Rightarrow_{\text{WSLD}} \langle G_n, \theta_n, \alpha_n \rangle$. And a WSLD refutation is a WSLD derivation $\langle G_0, \text{id}, 1 \rangle \Rightarrow_{\text{WSLD}} \langle \bot, \sigma, \alpha \rangle$, where $\bot$ is the empty clause, $(\sigma, \alpha)$ is a fuzzy computed answer, where $\sigma$ is a computed substitution and $\alpha$ its approximation degree. Certainly, a WSLD-refutation computes a family of answers, in the sense that, if $\theta = \{x_1/t_1, \ldots, x_n/t_n\}$ is a computed substitution, with degree $\alpha$, then any substitution $\theta' = \{x_1/s_1, \ldots, x_n/s_n\}$ satisfying $\mathcal{R}(s_i, t_i) \geq \lambda$, for any $1 \leq i \leq n$, is also a computed substitution if its approximation degree $\beta = \alpha \Delta(\Delta \gamma \mathcal{R}(s_i, t_i)) \geq \lambda$.

**E. Compilation of FuzzyDES Programs**

Similarly to Boussi-Prolog, FuzzyDES implements the operational mechanism of WSLD resolution by compiling programs into core Datalog rules (instead of to Prolog clauses). The idea is to obtain a set of Datalog rules of the form $\exists \mathcal{E}$. The compilation procedure, splitting it in two steps: i) the crisp unification of a defined predicate (in the head of the transformed rule); and ii) the weak unification of the attributes of the defined predicate (in the body of the transformed rules) with the help of the built-in weak unification predicate explained in Subsection IV-C. To this end, the transformation moves each head argument to an explicit weak unification performed by a specific predicate, as formalized later.

Moreover, if there exists $\mathcal{R}(p, q) = \alpha$ between predicates $p$ and $q$, then simulating a flexible matching of these predicate symbols by using a classical unification technique can be done by introducing a new clause for each predicate $q$ which is close to $p$.

The following definition formalizes the program transformation just outlined above, but first we need to introduce an extended language obtained by adding to the source language alphabet the elements of the lattice $[0,1]$ (of approximation degrees). Clauses in this extended language contain bodies with literals which are interpreted as approximation degrees. We call these clauses e-clauses (expanded clauses). Also, e-clauses with an empty head are called e-goals.

**Definition 4.2:** Let II be a logic program, $\mathcal{R}$ a similarity relation on the syntactic domain generated by II, $\Delta$ the fixed t-norm associated to $\mathcal{R}$, and $\lambda$ in $[0,1]$ a cut value. Let $\{p_1, \ldots, p_n\} \leftarrow Q; \delta$ be a graded clause in II defining the n-ary predicate $p$. Then, for each $\mathcal{R}(p, q) = \alpha > \lambda$ add to the transformed program $\Pi'$ the new e-clause:

$$q(x_1, \ldots, x_n) \leftarrow (\Delta \alpha \delta \alpha \theta x_1 \approx t_1 \land \cdots \land x_n \approx t_n \land Q)$$

where each $x_i$ is a fresh variable and $x_i \approx t_i$ is an expression that returns the approximation degree obtained from the weak unification of the term bound to the variable $x_i$ and the term $t_i$.

Observe that, since $\mathcal{R}(p, p) = 1$ for any symbol $p$, if $\{p_1, \ldots, p_n\} \leftarrow Q; \delta$ is in the original program, the e-clause $q(x_1, \ldots, x_n) \leftarrow \delta \land x_1 \approx t_1 \land \cdots \land x_n \approx t_n \land Q$ will be in the transformed program. Thus, we give a uniform treatment for all clauses in the transformed program. Since this transformation was initially implemented in Boussi-Prolog, we name it (for short) BPL expansion, and the transformed programs generated by its application BPL expanded programs. This transformation has been adapted and implemented in the FuzzyDES system with the predicate expand_rule/7 (called at compilation-time) and produces Datalog code able to be executed by the FuzzyDES system. The behaviour of this predicate is illustrated by the following example.

**Example 1:** Consider the following fragment of a simple FuzzyDES program, where the truth degree of a graded rule is specified with the operator with:

% PROXIMITY EQUATIONS
p1/q:i=0.6;
%

% FACTS AND RULES
r(a). p(b). p(X) :- r(X) with 0.8. q(c).

After the execution of the predicate expand_rule/7, the following Datalog code is generated ( having issued the command /fuzzy_expansion bpl):

r(A,D) :- '$unify_arguments'([[A,a,B]]), '$t_norm'([D,B],D), p(A,D) :- '$unify_arguments'([[A,b,B]]), '$t_norm'([D,B],D), p(A,D) :- '$unify_arguments'([[A,X,B]],r(X,C)), '$t_norm'([D,0.8,C,B],D), p(A,D) :- '$over_lambda_cut'(0.6), '$unify_arguments'([[A,c,B]]), '$t_norm'([D,B,0.6],D), q(A,D) :- '$unify_arguments'([[A,c,B]]), '$t_norm'([D,B,0.6],D), q(A,D) :- '$over_lambda_cut'(0.6), '$unify_arguments'([[A,b,B]]), '$t_norm'([D,B,0.6],D), q(A,D) :- '$over_lambda_cut'(0.6), '$unify_arguments'([[A,X,B]],r(X,C)), '$t_norm'([D,0.8,C,B,0.6],D).

Note that each transformed predicate has an additional argument in order to store and allow the propagation of truth degrees. Here, $\text{	exttt{$\$unify$arguments$}}$ implements the weak unification operator $\approx$ in Definition 4.2, and $\text{	exttt{$\$t_norm$}}$ (as explained in Subsection IV-B) is an internal predicate used to propagate the truth degrees of the rules and the approximation degrees coming from the fuzzy relations. The predicate

The rules listed as a result of the transformation correspond to core DATALOG, and they can be examined with the command /listing and enabling development listings with /development on.
$\texttt{\$over\_lambda\_cut}$ anticipates failure during solving of the expanded rule if the current $\lambda$-cut is above the approximation degree given by the equation corresponding to the expansion. Since $\text{p} / \text{q}$ is close to $\text{q} / \text{p}$ (with degree 0.6), one rule for $\text{q} / \text{p}$ is added for every rule defining $\text{p} / \text{q}$ (and vice versa).

Though users can inspect core Datalog by issuing the command \texttt{/development on}, the default mode is intended for reasoning in the source Datalog level, so that users manipulate both graded rules and proximity equations at the high-level source syntax (e.g., interactively adding or removing them).

In general, adding as many rules as those defining $\text{p}$ for a predicate $\text{q}$ close to $\text{p}$ with a truth degree $\alpha$ might develop a space issue for a high number of rules. Fortunately, thanks to the tabling-based operational mechanism of FuzzyDES, it is possible to simplify this transformation. So, $\text{q}$ can be defined as a straightforward way to $\text{p}$ with a truth degree $\alpha$ (see the second item of the next definition).

**Definition 4.3:** Let $\Pi$ be a logic program, $\mathcal{R}$ a similarity relation on the syntactic domain generated by $\Pi$, and $\lambda \in [0,1]$ a cut value. For each predicate $\text{p}$ defined in $\Pi$,  
1) for each clause $\text{p}(t_1, \ldots, t_n) \leftarrow Q, \delta$ in $\Pi$,  
   $\text{p}(x_1, \ldots, x_n) \leftarrow \delta \land x_1 \approx t_1 \land \cdots \land x_n \approx t_n \land Q$.
2) for each entry $\mathcal{R}(p, q) = \alpha \geq \lambda$ in $\mathcal{R}$ (with $\neq \text{p}$) add to the transformed program $\Pi'$ the e-clause:
   
   \[
   q(x_1, \ldots, x_n) \leftarrow \alpha \land p(x_1, \ldots, x_n).\]

   where each $x_i$ is a fresh variable.

Since this transformation is specific of the FuzzyDES system, we name it (for short) \texttt{FDES expansion}, and the transformed programs generated by its application \texttt{FDES expanded programs}.

Under this definition, if a program with $n$ predicates, each one defined by $r_i$ rules, and being close to other $c_i$ predicates, then last program transformation generates $\sum_{i=1}^{n} r_i + c_i$ rules. However, the program transformation of Definition 4.2 generates $\sum_{i=1}^{n} r_i + (r_i \times c_i)$ rules. Therefore, many transformed rules can be saved for a proximity (or similarity) relation with a huge amount of defined rules.

Note that the transformation of Definition 4.3 is only viable in the context of languages, such as Datalog, with specific syntactic constraints and using a tabling-based computation strategy. However, languages that follow a top-down strategy, like Bousi-Prolog, cannot benefit from this transformation. Note that this transformation can convert a program with a finite search space into one with an infinite search space for a purely top-down computational strategy (for instance, think of $\mathcal{R}(p, q) = \alpha_1$ and $\mathcal{R}(q, p) = \alpha_2$ that would lead to mutually recursive e-clauses).

The following example shows a transformation following Definition 4.3 by enabling it with the command \texttt{/fuzzy\_expansion on}.

**Example 2:** Consider again the program fragment of Example 1. After the execution of the predicate \texttt{expand\_rule7} the following Datalog code is generated:

\[
r(A, D) : = $\texttt{\$unify\_arguments}'([[A, a, B]]),$
\]

\[
p(A, D) : = $\texttt{\$unify\_arguments}'([[A, A, B]]),$
\]

\[
p(A, D) : = $\texttt{\$unify\_arguments}'([[A, X, B]]), r(X, C),$ \]

\[
p(A, D) : = $\texttt{\$unify\_arguments}'([[A, B, B]]),$ \]

\[
p(A, D) : = $\texttt{\$over\_lambda\_cut}'(0.6), p(A),$
\]

\[
p(A, D) : = $\texttt{\$t\_norm}'('[[A, B]], D).$
\]

\[
p(A, D) : = $\texttt{\$t\_norm}'('[[A, B]], D).$
\]

\[
p(A, D) : = $\texttt{\$t\_norm}'('[[A, B]], D).$
\]

\[
p(A, D) : = $\texttt{\$t\_norm}'('[[A, B]], D).$
\]

\[
p(A, D) : = $\texttt{\$t\_norm}'('[[A, B]], D).$
\]

\[
q(A, D) : = $\texttt{\$over\_lambda\_cut}'(0.6), p(A),$ \]

\[
q(A, D) : = $\texttt{\$over\_lambda\_cut}'(0.6), p(A),$ \]

\[
q(A, D) : = $\texttt{\$over\_lambda\_cut}'(0.6), p(A),$ \]

\[
solving e-clauses from this transformation in the deductive setting poses no non-termination problems.

Finally, as we shall justify in Section V, the Datalog code generated by the transformation of either Definition 4.2 or 4.3 (depending on the flag \texttt{fuzzy\_expansion}) is executed by the deductive engine of the FuzzyDES system, therefore emulating the result of executing the original program under WSLD resolution.

**F. Fuzzy Answer Subsumption**

As introduced in Subsection III-D, answer subsumption in a deductive database simply resorts to a membership test. Indeed, this approach can be also used in the fuzzy setting, but a different, more-efficient notion of answer subsumption can be applied as we propose at the end of this subsection.

Following any of the expansions explained in the former Subsection IV-E, each program rule is compiled to a rule with an additional argument for the approximation degree. FuzzyDES handles a user query $Q$ by transforming the query into an \texttt{autoview} for the predicate \texttt{answer}\texttt{m}+1, where its body is the compiled query, and $m$ are the relevant variables in $Q$. The last argument of the transformed call is the placeholder for the approximation degree. The outcome to the user query is thus built by solving the call $\text{answer}(X_1, \ldots, X_m, X_\delta)$, where $X_1, \ldots, X_m \subseteq \text{vars}(Q)$ are the relevant variables in the user query, and $X_\delta$ is the variable for the approximation degree. Each matching tuple in the answer table (i.e., $t_1 = \text{answer}(X_1, \ldots, X_m, X_\delta)$) is presented to the user as $\text{answer}(X_1, \ldots, X_m, X_\delta)$ with $X_\delta$. The answer table holds entries for all user predicates involved in solving a user query, each one with the last argument being the approximation degree, which is ground as the other arguments are. Then, along solving, it is possible to deduce a tuple $t_1 \equiv t(C, D)$ so that there already exists another tuple $t_2 \equiv t(C, D')$ in the answer table, differing only in the approximation degree. If $D' > D$, it is not worthwhile to add the same tuple with a lower approximation degree, so that we say that $t_2$ subsumes $t_1$. This can be seen as a generalization of answer subsumption to the fuzzy setting, and we apply in the deductive setting obtaining at least three advantages: First, the size of the answer table is reduced. Second, answer table look-ups are therefore more efficient. And, third, joins are simplified by taking into account less tuples.

**V. OPERATIONAL SEMANTICS FOR EXPANDED PROGRAMS**

This section formally describes, at a high abstraction level, the operational semantics for expanded programs implemented both by Bousi-Prolog and the FuzzyDES systems. This semantics simulates the WSLD resolution rule of Definition 4.1,
and is the basis to the implementation described in Subsection IV-E. The main goal of this section is to establish that, in fact, the behavior of a program executed by using WS LD resolution is equivalent to one of the corresponding expanded program when it is executed by this abstract operational mechanisms.

In the remainder of this section we shall work inside the framework of the extended language built by e-clauses and e-goals. II′ denotes a transformed program (or expanded program) obtained by applying either Definition 4.2 or Definition 4.3 on a logic program II equipped with a similarity relation \( R \) and a cut value \( \lambda \). In what follows, transition steps are applied to underlined fragments, and the symbols \( Q, Q' \) are denoting conjuncts of atoms and, possibly, approximation degrees.

**Definition 5.1:** We define the operational semantics for expanded programs as a transition system \( (E, \Rightarrow_E) \) where \( E \) is a set of triples \( \langle \{G, \theta, \alpha\} \) (e-goal, substitution, approximation degree), and the transition relation \( \Rightarrow_E \subseteq (E \times E) \) is the smallest relation which satisfies:

**Rule 1:** if \( \beta \in (0, 1] \) and \( (\beta \Delta \alpha) + \geq \lambda \)

\[ \langle \{\langle\beta \land Q', \theta, \alpha\rangle \Rightarrow \langle Q', \beta \Delta \alpha\rangle \rangle \]

**Rule 2:** if \( \sigma = \text{wmgd}_i^\beta (A, B) \neq \text{fail} \), \( \beta = \mathcal{R}(A \sigma, B \sigma) \geq \lambda \), and \( (\beta \Delta \alpha) + \geq \lambda \)

\[ \langle \{\langle\beta \equiv B \land Q', \theta, \alpha\rangle \Rightarrow \langle Q', \theta, \beta \Delta \alpha\rangle \rangle \]

**Rule 3:** if \( (p(x_1, \ldots, x_n) \leftarrow \beta \land x_1 \equiv t_1 \land \cdots \land x_n \equiv t_n \land Q) \) \( \not\in \Pi' \),

\[ \langle \{\langle p(s_1, \ldots, s_n) \land Q', \theta, \alpha\rangle \Rightarrow \langle \beta \equiv s_1 \equiv t_1 \land \cdots \land s_n \equiv t_n \land \theta, \alpha\rangle \rangle \]

Note that in the operational step defined by Rule 3, we perform a syntactic unification of the selected atom of the e-goal and the head of the e-clause. Note also that in this case, the most general unifier \( \{x_1/s_1, \ldots, x_n/s_n\} \) does not participate in the final computed answer because its domain variables are standardized apart (i.e., they are fresh variables) and does not affect the bindings of the substitution \( \theta \). In what follows, we will often speak of simplification steps to refer to the steps performed with Rule 1 above.

To achieve a more condensed notation in the proofs, throughout the remainder of this section, we introduce the following notation: we write \( \pi \) for the sequence of syntactic objects \( o_1, \ldots, o_n \); similarly, \( \pi \) denotes the composition of substitutions \( \sigma_1 \sigma_2 \cdots \sigma_n \).

A. Semantic Equivalence for BPL Expanded Programs

In the sequel, we prove the semantic equivalence between the WS LD rule, applied to a logic program II, and the operational mechanism of Definition 5.1, when it is applied to the BPL expanded program \( \Pi' \) (obtained from II and generated by Definition 4.2).

First we recall a useful property of a similarity relation \( R \) proved by M. Sessa.

**Proposition 5.2:** [7, pag. 397] Let \( R \) be a similarity relation and \( \lambda > 0 \) a cut value. For any substitution \( \theta \) and terms \( t, t' \), if \( R(t, t') \geq \lambda \), then \( R(t\theta, t'\theta) = R(t, t') \geq \lambda \).

**Lemma 5.3:** Given a program II with a similarity relation \( R \), an associated t-norm \( \alpha \), and a cut value \( \lambda \in (0, 1] \), let \( \Pi' \) be the BPL expanded program. If there is a step \( S' \):

\[ (\langle \langle \neg Q \land Q', \theta, \alpha \rangle \Rightarrow \langle \neg Q \land Q', \theta, \alpha \rangle \rangle \]

in \( \Pi' \), then there is a derivation: \( \langle \langle Q \neg Q' \rangle \rangle \Rightarrow \langle Q \neg Q' \rangle \rangle \) in \( \Pi' \), which computes the same state. **Proof.** If there is a step \( S \) in \( \Pi' \), is because there exists a graded rule \( C = \langle (p(t_n), \theta) \rangle \) in \( \Pi' \) such that

\[ \sigma = \text{wmgd}_i^\beta (q(s_n), p(t_n)) = \text{wmgd}_i^\beta (\langle s_n \equiv t_n \rangle \),

with approximation degree

\[ \gamma \begin{array}{c} = \mathcal{R}(q(s_n, \sigma, p(t_n))) \equiv \mathcal{R}(q, p) (\Delta^\alpha_n \cap \Delta^\alpha_n \cap \Delta^\alpha_n \geq \lambda), \end{array} \]

\[ \lambda \geq (\Delta^\alpha_n \lambda) \geq \lambda \]

On the other hand, by Definition 4.2, if \( R(p, q) = \beta \geq \lambda \), there is an e-clause \( q(s_n) \begin{array}{c} \leftarrow (\Delta^\alpha_n \cap \Delta^\alpha_n \cap \Delta^\alpha_n \geq \lambda), \end{array} \)

Therefore, it is possible to construct the following derivation in \( \Pi' \):

\[ (\langle \langle Q \neg Q' \rangle \rangle \equiv \langle \langle Q \neg Q' \rangle \rangle \rangle \]

\[ \Rightarrow \langle \langle \langle Q \neg Q' \rangle \rangle \equiv \langle \langle Q \neg Q' \rangle \rangle \rangle \rangle \]

\[ \Rightarrow \langle \langle \langle Q \neg Q' \rangle \rangle \equiv \langle \langle Q \neg Q' \rangle \rangle \rangle \rangle \]

\[ \Rightarrow \langle \langle \langle Q \neg Q' \rangle \rangle \equiv \langle \langle Q \neg Q' \rangle \rangle \rangle \rangle \]

where \( \sigma_i = \text{wmgd}_i^\beta (s_i \equiv t_i) \), for \( 2 \leq i \leq n \), \( \sigma_i = \text{wmgd}_i^\beta (s_i \equiv t_i) \), and, thanks to Proposition 5.2, \( \mathcal{R}(s_i, \sigma_i, t_i, \sigma_j) = \mathcal{R}(t_i, s_i, \sigma_j) = \lambda_i \)

Note that, when we apply the weak unification algorithm described in Section IV-A to the unification problem \( \{s_n \equiv t_n\} \), we reach a successful configuration:

\[ (\langle s_1 \equiv t_1, \ldots, s_n \equiv t_n \rangle, id, 1) \]

\[ \Rightarrow (\langle s_2 \equiv t_2, \ldots, s_n \equiv t_n \rangle, \sigma_1, \lambda_1') \]

\[ \Rightarrow (\langle s_3 \equiv t_3, \ldots, s_n \equiv t_n \rangle, \sigma_2, \lambda_1', \lambda_2') \]

\[ \cdots \]

\[ \Rightarrow (\langle s_n \equiv t_n \rangle, \sigma_n, \lambda_1', \cdots, \lambda_1') \]

and this weak unification process closely follows the steps performed with the subgoals \( s_i \equiv t_i \) in the former derivation with the CPL expanded program. Therefore, \( \sigma_n = \text{wmgd}_i^\beta (\langle s_n \equiv t_n \rangle) = \sigma \) and \( R(\langle s_n \equiv t_n \rangle) = \Delta^\alpha_n \lambda_i \).

The following proposition establishes a kind of completeness result where we prove that derivations in the original program using the WS LD resolution rule can be reproduced by the FuzzyDES operational mechanism in the transformed program.

**Proposition 5.4:** Given a program II with a similarity relation \( R \), let \( \Pi' \) be the BPL expanded program. If there
is a derivation $D = \langle \leftarrow Q, \theta, \alpha \rangle \Rightarrow_{\text{WSLD}}^* \langle \leftarrow Q', \theta', \alpha' \rangle \rangle$ in $\Pi$, then there is a derivation $\langle \leftarrow Q, \theta, \alpha \rangle \Rightarrow_{\text{Ex}}^* \langle \leftarrow Q', \theta', \alpha' \rangle$ in $\Pi'$, which computes the same state.

**Proof.** By induction on the length of the derivation $D$ and Lemma 5.3.

Now we proceed by demonstrating the reverse of the last proposition, which constitute a kind of soundness result.

**Lemma 5.5:** Given a program $P$ with a similarity relation $R$, an associated t-norm $\Delta$, and a cut value $\lambda \in \{0,1\}$, let $\Pi'$ be the BPL expanded program. If there is a derivation $D' = \langle \leftarrow p(\pi_n) \wedge Q', \theta, \alpha \rangle \Rightarrow_{\text{Ex}}^+ \langle \leftarrow (Q \wedge Q')\sigma, \theta, \lambda \Delta \alpha \rangle$ in $\Pi'$, then there is a step $S: \langle \leftarrow p(\pi_n) \wedge Q', \theta, \alpha \rangle \Rightarrow_{\text{WSLD}}^* \langle \leftarrow (Q \wedge Q')\sigma, \theta, \lambda \Delta \alpha \rangle$ in $\Pi$.

**Proof.** Note that we can assume that the shape of the derivation $D'$ is:

$$
\langle \leftarrow (Q \wedge Q')\sigma, \theta, \lambda \Delta \alpha \rangle
$$

where the first step is performed by Rule 3 and then it is followed by a sequence of applications of Rule 2, with $\lambda_i = R(t_1, s_1, \sigma) \wedge \ldots \wedge R(t_n, s_n, \sigma)$. Finally, a simplification step with Rule 1 is performed and the degree $\beta$ is compounded with $\lambda_i = \lambda'$.

If the first step of derivation $D'$ is possible, it is because there exists an e-clause $C' = \langle \leftarrow (\delta \Delta \beta) \wedge x = t_n \wedge Q \rangle$ in $\Pi'$ and there must be an entry $R(q, p) = \beta \geq \lambda$ in $\Pi$. So, there exists a clause $C = \langle \leftarrow q(t_n) \wedge Q, \sigma \rangle$ in $\Pi$, whose head weakly unify with $p(\pi_n)$ and $\text{wmgu}_N^k(t_n, p(\pi_n)) = \sigma$ with approximation degree $\text{R}(q(t_n), \sigma, p(\pi_n)) = \text{R}(q(t_n), \sigma, p(\pi_n)) = \beta \Delta (\Delta_{i=1}^n \lambda_i) = \lambda'$. Therefore, it is possible the WSLD step $S$.

**Proposition 5.6:** Given a program $P$ with a similarity relation $R$, let $\Pi'$ be the BPL expanded program. If there exists a derivation $D' = \langle \leftarrow Q, \theta, \alpha \rangle \Rightarrow_{\text{Ex}}^* \langle \leftarrow Q', \theta', \alpha' \rangle \rangle$ in $\Pi'$, then there exists a derivation $\langle \leftarrow Q, \theta, \alpha \rangle \Rightarrow_{\text{WSLD}}^* \langle \leftarrow Q', \theta', \alpha' \rangle$ in $\Pi$, which computes the same state.

**Proof.** By induction on the length of the derivation $D'$. Without loss of generality we can assume that the steps in derivation $D'$ are conveniently ordered to allow the application of Lemma 5.5.

The last proposition jointly with Proposition 5.4 state the equivalence of both operational mechanisms and the correctness of our implementation.

**B. Semantic Equivalence for FDES Expanded Programs**

In this subsection we turn our attention to FDES expanded programs. The ultimate objective, as in the previous subsection, is to prove the equivalence of WSLD resolution and the source program $P$ with respect to the operational mechanism of Definition 5.1. An application to an FDES expanded program $\Pi'$ obtained from $P$. In this case, we proceed in an indirect way, studying the relation between these two types of transformed programs, and proving that they are semantically equivalent (modulo subsumed answers) with respect to the operational semantics for expanded programs of Definition 5.1.

Given a source program $P$, we first prove that the BPL expanded program $\Pi'$ for $P$ results from the unfolding transformation of the corresponding FDES expanded program $\Pi''$.

**Program transformation** is an optimization technique for computer programs that, from an initial program $\Pi_0$, derives a sequence $\Pi_1, \ldots, \Pi_n$ of transformed programs by applying **elementary transformation rules** which improve the original program under some criteria.

**Unfolding** is a well-known semantics-preserving program transformation rule (first introduced in [27] to optimize functional programs). In essence, it is usually based on the application of operational steps on the body of program rules. The unfolding transformation is able to improve programs, generating more efficient code. Unfolding is the basis for developing sophisticated and powerful programming tools, such as folding/unfolding transformation systems and partial evaluators.

In our fuzzy framework, the unfolding transformation can be defined, similarly to [28], as follows:

**Definition 5.7:** Let $\Pi''$ be an expanded program and $R : (A \leftarrow Q) \in \Pi$ a (non unit) program rule. Then, the fuzzy unfolding of $\Pi''$ with respect to the rule $R$ and a fixed selection rule is the new expanded program $\Pi'' = (\Pi'' \cup U) \setminus \{R\}$ such that:

$$
U = \{ A \sigma \leftarrow \alpha \wedge Q' \mid \langle Q; id; 1 \rangle \Rightarrow_{\text{Ex}} \langle Q'; \sigma; \alpha \rangle \}
$$

In order to accelerate the unfolding transformation process, we will allow to perform a sequence of simplification steps after an unfolding step.

A BPL expanded program can be obtained by unfolding the FDES expanded program. Next proposition states this result.

**Proposition 5.8:** Given a program $P$ with a similarity relation $R$, an associated t-norm $\Delta$, and a cut value $\lambda \in \{0,1\}$, let $\Pi''$ be the FDES expanded program. The BPL expanded program $\Pi''$ can be obtained by unfolding $\Pi'$, after disregarding rules that produce subsumed answers.

**Proof.** Without loss of generality, we can assume a program $P$ defining two predicates $p$ and $q$, each one defined by $k$ and $l$ rules respectively, and equipped with a similarity relation $R$ (characterized by the reflexive, symmetric, transitive t-closure of the entries $R(p, q) = \alpha, R(q, r) = \beta$ and $R(p, r) = \gamma$).

Note that since the predicate $r$ is not defined by rules, there are no e-clauses for it.

$$
\Pi'' = \{ R_{p_1} : p(\pi_n) \leftarrow Q_1, \delta_{p_1} ; \ldots, R_{p_k} : p(t_{kn}) \leftarrow Q_k, \delta_{p_k} , R_{q_1} : q(\pi_n) \leftarrow B_1, \delta_{q_1} ; \ldots, R_{q_l} : q(t_{ln}) \leftarrow B_1, \delta_{q_l} \}
$$

For this program, the FDES expanded program is:

$$
\Pi^* = \{ R_{p_1}^1 : p(\pi_n) \leftarrow \delta_{p_1} \wedge \pi_n \approx t_{kn} \wedge Q_1, \ldots, R_{p_1}^{\lambda} : p(t_{kn}) \leftarrow \delta_{p_1} \wedge t_{kn} \approx Q_k, R_{p,q} : q(\pi_n) \leftarrow \alpha \wedge p(\pi_n), R_{p,r} : r(\pi_n) \leftarrow \gamma \wedge p(\pi_n), \}
$$

Copyright (c) 2018 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.
Now, we proceed to unfold $\Pi^1$ with respect to $R_{p,q}$ using a selection rule that selects the subgoal $p(x_n)$, in the body of $R_{p,q}$. Then, for each rule $R^1_{pi}$ in $\Pi^1$:

$$R^1_{q_1} : q(x_n) \leftarrow \delta_{q_1} \land x_n \approx s_{1n} \land B_1,$$

$$\ldots$$

$$R^1_{q_k} : q(x_n) \leftarrow \delta_{q_k} \land x_n \approx s_{kn} \land B_1,$$

$$R_{q,p} : p(x_n) \leftarrow \alpha \land q(x_n),$$

$$R_{q,r} : r(x_n) \leftarrow \beta \land q(x_n)$$

leading, by Definition 5.7, to the rules:

$$q(x_n) \leftarrow (\alpha \land q(x_n))$$

For the rule $R_{q,p}$, also defining $p$, we have the derivation

$$\langle \alpha \land q(x_n), id, 1 \rangle \xrightarrow{R_{q,p}} \langle \alpha \land q(x_n), id, 1 \rangle$$

This last derivation leads to the recursive rule

$$q(x_n) \leftarrow (\alpha \land q(x_n))$$

leading to a sequence of unfolded programs:

$$\Pi^1 = \{ R^2_{q_1} : q(x_n) \leftarrow (\alpha \land q(x_n)),$$

$$\ldots$$

$$R^2_{q_k} : q(x_n) \leftarrow (\alpha \land q(x_n)) \}$$

with this first unfolding step we obtain an unfolded program

$$\Pi^2 = (\Pi^1 \cup \Pi^1) \setminus \{ R_{p,q} \}.$$ 

Note that the rule $R_{p,q}$ in $\Pi^1$ has been removed and the rules of the BPL expanded program produced by the $k$ rules defining $p$ in $\Pi$ and the entry $\mathcal{R}(p,r) = \gamma$ (see Definition 4.2) have been added.

In a complete similar way we can unfold $\Pi^3$ w.r.t to $R_{q,p}$ given the unfolded program $\Pi^4$ and in a further unfolding transformation we can unfold $\Pi^4$ w.r.t to $R_{q,r}$ given the unfolded program $\Pi^5$ where the rules $R_{q,p}$ and $R_{q,r}$ have been removed and the rules of the BPL expanded program produced by the $l$ rules defining $q$ in $\Pi$ and the entries $\mathcal{R}(q,p) = \alpha$ and $\mathcal{R}(q,r) = \beta$ (see Definition 4.2) have been added. As a result $\Pi^5$ is the BPL expanded program.

In summary, starting from the FDES expanded program, $\Pi^1$, we carry out continuous unfolding transformation steps, leading to a sequence of unfolded programs: $\Pi^2, \Pi^3, \Pi^4, \Pi^5$, which ends into the BPL expanded program for $\Pi$.

Finally, we can prove that an FDES expanded program executed by the operational mechanism of Definition 5.1 is semantically equivalent to the original program executed by WSLD resolution.

**Proposition 5.9:** Given a program $\Pi$ with a similarity relation $\mathcal{R}$ and a cut value $\lambda \in (0,1]$, let $\Pi'$ be the FDES expanded program. $\Pi'$, executed by the operational mechanism of Definition 5.1, is semantically equivalent to $\Pi$ executed by WSLD resolution. That is, they produce the same fuzzy computed answers (modulo subsumed answers).

**Proof.** Given a program $\Pi$ and its corresponding FDES expanded program $\Pi'$, by Proposition 5.8, the BPL expanded program $\Pi''$ is the unfolded program obtained from $\Pi'$ if some rules that produce subsumed answers are removed. First note that, unfolding is a semantics-preserving program transformation (preserving fuzzy computed answers). On the other hand, according to propositions 5.6 and 5.4, the BPL expanded program $\Pi''$, executed by the operational mechanism of Definition 5.1, is semantically equivalent to the original program executed by WSLD resolution. Therefore, we can affirm that an FDES expanded program executed by the operational mechanism of Definition 5.1 is semantically equivalent to the original program executed by WSLD resolution (modulo subsumed answers).

**VI. PERFORMANCE**

This section tries to highlight the effect of different parameters of the system with respect to their impact on its performance. Solving time and data structure sizes are elements of interest in performing experiments. A number of tests have been executed with different parameters to analyse scalability. As test platform, we used a Windows 10 64 bit OS on an Intel Xeon CPU E3-1505M v5 (4 physical cores) running at 2.8 GHz at its peak, with 16GB RAM. DES (programmed in Prolog) is run with SICStus Prolog 4.3.1 64 bit in interpreted mode (no compiled executable). Solving time is only due to solving, eliding e.g. parsing and display. Statistics in DES are enabled to collect data as timings, look-ups, and data structure sizes. This incurs in a small but noticeable overhead on the
system performance. For the fuzzy environment, we consider a similarity (transitive) relation \( \sim \) with a product t-norm.

Regarding solving parameters, we analyse the impact of the two different expansions BPL and FDES. The next two subsections focus on this: the first one on extensional predicate similarity, and the second one on intensional predicate similarity. We do not handle the case of constant similarity because it does not affect expansions, so that the same performance results would be retrieved for both. Subsection VI-C analyses the performance impact of enabling answer subsumption for both expansions. Finally, Subsection VI-D analyses the performance impact of specifying a \( \lambda \)-cut for both expansions.

### A. Comparing Expansions for Extensional Predicate Similarity

A first experiment consists of examining the performance of the two expansions BPL and FDES for extensional predicates (consisting only of facts). Here, we consider \( n \) extensional predicates so that they are similar as follows: \( p_{i}/2 \sim p_{i+1}/2 = 0.5 \) for \( i \in \{1, \ldots, n-1\} \). Each predicate is defined by \( m \) facts of the form: \( p_i(c_{i+1}, \ldots, c_{i+j-1}) \), where \( i \in \{1, \ldots, n\} \) and \( j \in \{1, \ldots, m\} \). For example, for \( n = 2 \) and \( m = 3 \), the resulting test program is:

\[
\begin{align*}
p1/2 & \sim p2/2 \sim 0.5, \\
p1(c1, c2) & \sim p2(c1, c2), p1(c2, c3), p1(c3, c4), \\
p2(c2, c4) & \sim p2(c3, c5), p2(c4, c6).
\end{align*}
\]

This way, a pair of constants which is not in a predicate \( p_i \) but is in \( p_i \), can be deduced for \( p_i \) with an approximation degree less than one, and as small as the product distance between \( p_i \) and \( p_i \). As parameter instances we selected \( n \in \{2, 4, 6\} \), and \( m \in \{10, 100, 200, 500, 1000, 2000, 5000, 10000, 15000, 20000\} \). A failing goal \( p1(a, b) \) (which produces no answer tuples) has been selected with the aim to traverse all the rules in the database (instead of testing the performance of caching due to tabling), either directly to the predicate \( p/2 \) or indirectly to the rest of the predicates via the similarity relation.

Table I includes results for these parameters \( m \) and \( n \), and for each combination of them, three measures are shown: Columns “So.”, “Co.”, “Rs.” respectively show the ratio of solving time, consult time, and number of rules. Ratios are always shown as the BPL measure with respect to the FDES measure. For each test configuration, 10 runs have been executed, and the best one has been selected.

### B. Comparing Expansions for Intensional Predicate Similarity

Following the comparison of both expansions, a second experiment considers a database composed of both extensional and intensional predicates (with non-empty rule bodies). We use the same parameters \( n \) and \( m \) as in last subsection, and introduce a new parameter \( k \). By contrast to the former experiment, here we focus on similarities between \( k \) intensional predicates out of \( n \). We denote the \( k \) intensional predicates as \( p_1, \ldots, p_k \), while the \( n-k \) extensional predicates are denoted as \( p_{k+1}, \ldots, p_n \). The form of each intensional predicate clause is \( p_i(X_1, X_{n-k+i}) \sim p_{k+i+1}(X_1, X_2), p_{k+i+2}(X_2, X_3), \ldots, p_n(X_{n-k+i}, X_{n-k+i}) \), where \( i \in \{1, \ldots, k\} \).

Hence, intensional predicate bodies are defined as calls to each extensional predicate, correlated with shared variables so that a call \( p_{k+i+1}(X_1, X_{i+1}) \) is followed by \( p_{k+i+1}(X_1, X_{i+2}) \) for \( i < n - k \). As for the intensional predicate head, the first (resp. second) head variable is the first (resp. last) variable in the body. As it was just commented, similarity is only specified

### Table I

**Extensional Predicate Similarity Expansions Comparison**

<table>
<thead>
<tr>
<th></th>
<th>( m = 2 )</th>
<th>( m = 4 )</th>
<th>( m = 6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.07</td>
<td>1.08</td>
<td>1.62</td>
</tr>
<tr>
<td>100</td>
<td>1.36</td>
<td>1.20</td>
<td>1.95</td>
</tr>
<tr>
<td>200</td>
<td>0.86</td>
<td>1.15</td>
<td>1.97</td>
</tr>
<tr>
<td>500</td>
<td>1.13</td>
<td>1.31</td>
<td>1.99</td>
</tr>
<tr>
<td>1,000</td>
<td>1.00</td>
<td>1.54</td>
<td>1.99</td>
</tr>
<tr>
<td>2,000</td>
<td>1.14</td>
<td>1.88</td>
<td>2.00</td>
</tr>
<tr>
<td>5,000</td>
<td>1.31</td>
<td>2.74</td>
<td>2.00</td>
</tr>
<tr>
<td>10,000</td>
<td>1.32</td>
<td>3.25</td>
<td>2.00</td>
</tr>
<tr>
<td>15,000</td>
<td>1.38</td>
<td>3.18</td>
<td>2.00</td>
</tr>
<tr>
<td>20,000</td>
<td>1.39</td>
<td>3.48</td>
<td>2.00</td>
</tr>
</tbody>
</table>
for intensional predicates. We assign, axiomatically, proximity equations for the \( k \) intensional predicates. Specifically, we write \( p_i/2 \sim p_{i+1}/2 = 0.5 \) for \( i \in \{1, \ldots, k-1\} \). On the other hand, the rest of \( n-k \) (extensional) predicates are defined as in the first experiment, i.e., each one defined with \( m \) facts. For example, for \( n=4, m=2, k=2 \), the resulting test program is:

\[
\begin{align*}
\text{p1}(X1, X3) & :- \text{p3}(X1, X2), \text{p4}(X2, X3). \\
\text{p2}(X1, X3) & :- \text{p3}(X1, X2), \text{p4}(X2, X3). \\
\text{p3}(c1, c2). & \text{p4}(c2, c3). \\
\text{p4}(c1, c2). & \text{p4}(c2, c4).
\end{align*}
\]

We consider for the experiments 10 predicates \( (n = 10) \), \( m \in \{10, 100, 200, 500, 1000, 2000, 5000, 10000, 15000, 20000\} \), and \( k \in \{3, 5, 7\} \). The same goal \( \text{p1}(a, b) \) has been considered.

Table II includes results for these parameters \( m \) and \( k \), and for each combination of them, two measures are shown: “So.” and “Co.” with the same meaning as before but with the ratios the other way round. Observe that the number of rules in the expansion coincides for both BPL and FDES (we therefore omit the column “Rs.”), and ranges from 87 to 140,017 (including the rules for t-closure computation).

As in the first experiment, fluctuation do occur for small solving times (less than 50 ms) related to small numbers for \( m \). In this case, focusing on larger numbers of \( m \), solving times are better in BPL than in FDES, and increase as \( m \) grows. It is also observed that higher speed-ups are got for larger \( k \) (number of intensional predicates). Though there are the same expanded rules in both cases, in FDES, there is one mutually recursive rule for each intensional predicate instead of a non-recursive rule in the case of BPL. Because (mutually) recursive rules are harder to solve in DES as already mentioned, this issue explains such behaviour. With respect to consult times, both expansions converge to 1 for larger numbers of \( m \) because the same number of rules are generated. However, for small numbers of \( m \), there is a timing threshold due to PDG construction, which follows a naïve algorithm. Again, recursive predicates make this algorithm to perform badly. This is also the reason for omitting PDG updating when submitting queries, which are automatic views (therefore rules added to the database and hence, the PDG should be updated). For the queries posed, there is no need for PDG update since in particular the stratification does not change.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( m )</th>
<th>( k )</th>
<th>( l )</th>
<th>Exp.</th>
<th>S. S-U</th>
<th>ET. Gain</th>
<th>Ans. Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10</td>
<td>2</td>
<td>4</td>
<td>BPL</td>
<td>1.078</td>
<td>365</td>
<td>1.025</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>4</td>
<td>10</td>
<td>BPL</td>
<td>1.2</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

C. Analysing Answer Subsumption

Subsection IV-F explained the concept of fuzzy answer subsumption to prune computations as an optimization to the naïve case. Here we test the benefits of this optimization with respect to both solving time and data size. By contrast to former experiments, the goal must succeed and return answer tuples to examine the amount of such tuples, the tuples in the extension table, and its solving time. Here we select the goal \( \text{p1}(c1, x) \). Regarding the test program, we reuse the one in last subsection, i.e., each one defined with \( m \) facts. For the next value \( (0.05, very close to 0.00) \) a great improvement is achieved (even when answer subsumption is enabled) since many computations derived from constant

\( \lambda \)-cut

Finally, this subsection analyses the effect of the \( \lambda \)-cut value for both expansions. Again, we consider the same benchmark and parameters as in the last subsection with \( (n, m, k, l) = (7, 100, 4, 30) \), that is, we specify 4 similar intensional predicates (out of 7, with 3 extensional predicates each one with 100 facts) and 30 similar constants. This amount of similar constants will highlight the effect of the \( \lambda \)-cut for both expansions. The goal is the same as in the last subsection.

Table IV includes the column \( \lambda \)-cut for different selected values, showing results for both expansions BPL and FDES and both for solving time and ET size gains. These columns have the same meaning as in the last subsection.

Measures for the value 0.00 of \( \lambda \)-cut correspond to the worst case and are therefore compared to itself (ratios of 1). For the next value (0.05, very close to 0.00) a great improvement is achieved (even when answer subsumption is enabled) since many computations derived from constant
similarity are pruned. Note that in all cases, the number of answers is the same and, thus, the column "Ans. Gain" is elided. This improvement increases as \( \lambda \)-cut does. For the last value, there is no further computation cuts and the gains remain constant. It is also worth to note that the achievements are better in FDES than in BPL because more computations (due to the recursive part) are cut. Though, the computation time is still better in BPL.

VII. RELATED WORK

The idea of integrating concepts derived from fuzzy logic within a database is not new. In this section we give an abbreviated record of these achievements (see the supplementary material provided at IEEE Xplore for a more detailed study on fuzzy databases).

Buckles and Petry [29] fuzzify relational databases by associating similarity relations to the attribute domains in a database. In that way, they can weaken the equality relation in order to perform some kind of flexible querying through a extended relational algebra. Later on, Shenoi and Melton [30] extended this model to deal with proximity relations. Note that, although we share with these authors the use of similarity/proximity relations, the methods they use to represent and deal with the fuzzy component of a database differ by far from ours.

Umano [31] proposed an alternate model supported on the concept of possibility distribution. It was generalized by Prade and Testemaele [32] and it turned into a mainstream in the area. FREDDI [10] synthesizes the most outstanding aspects of this prior approach but incorporating deductive capabilities. An extended implementation of the FREDDI architecture [11] adopts a clausal representation for the rules of the intensional database, and deductive capabilities are obtained by implementing a non-classical bottom-up algorithm inspired in Datalog techniques. Although this framework presents deductive characteristics, unlike ours, it requires the user, explicitly, to take into account the propagation of the truth degrees through an additional parameter that stores a “matching” degree, leading to a kind of hybrid system combining SQL and logic rules. In addition, though they claim that their proposal is based on two theoretical models, there is no theoretical provision for their combination. Moreover, this system does not seem to be available.

The introduction of fuzzy features in classical deductive databases, with Datalog as a query language, is a more recent activity and has not produced so many examples. In this field, some of the work has consisted in the addition of confidence factors or truth degrees to the facts and rules of the extensional and the intensional databases which are manipulated by fuzzy connectives (e.g., the work presented in [33]). However, we are interested in those works that introduce fuzzy characteristics within the framework of the Datalog language, primarily through the use of fuzzy relations. In this sense, the work of A. Achs et al. [9] can be considered as a pioneering work, since it integrated the use of fuzzy relations in the framework of the Datalog language for the first time. They started from an extended Datalog language, with truth degrees in facts and rules, to which they incorporated a similarity relation on the domain of attributes. Unfortunately, and contrary to our proposal, this work has the following limitations: i) it only considers the similarity between the attributes of a relation, not between the relation itself; and ii) Datalog programs undergo a transformation guided by the similarity relation, but this transformation is complex (and possibly inefficient) and does not adequately handle rules with non-variable arguments. Finally, to the best of our knowledge, there is no report of a publicly available implementation of their system.

The work of A. Achs et al. was developed in parallel and independently with respect to the works [34], [35] and [7]. There, they introduced similarity relations in a logic programming language, and the concept of unification by similarity was first developed. It is in this last stream, and especially in the works of M. Sessa, that we have found our inspiration. As within the Bousi-\( \sim \)Prolog language [5], [6], in the FuzzyDES system we make use of the unification algorithm proposed in [7]. However, there are some striking differences between the concepts presented in [7] and those we use:

1) We adopt a different notion of weak unifier and weak most general unifier. ii) Also, we have defined a more appropriate concept of similar clause, which has a beneficial impact on the efficiency of the system we have implemented. iii) We have introduced some new peculiarities into the operational semantics of our language (such as the use of a threshold or a different notion of family of computed answers). iv) Finally, we have developed new implementation techniques for efficient execution of programs. These conceptual and practical differences must be added to the differences arising from the fact that FuzzyDES uses an operational mechanism based on tabling (a bottom-up top-down-driven strategy) which have been adequately explained throughout this work.

VIII. CONCLUSIONS

This paper has introduced formal and practical aspects needed to transfer some techniques found in the fuzzy logic programming system Bousi-\( \sim \)Prolog into the deductive system DES. The new extension of DES, called FuzzyDES, defines a fuzzy Datalog language which is executed by the operational

---

<table>
<thead>
<tr>
<th>( \lambda )-cut</th>
<th>Exp.</th>
<th>S. S-U</th>
<th>ET. Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>BPL</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>DES</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.05</td>
<td>BPL</td>
<td>93</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>DES</td>
<td>120</td>
<td>11</td>
</tr>
<tr>
<td>0.10</td>
<td>BPL</td>
<td>146</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>DES</td>
<td>190</td>
<td>16</td>
</tr>
<tr>
<td>0.25</td>
<td>BPL</td>
<td>247</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>DES</td>
<td>316</td>
<td>27</td>
</tr>
<tr>
<td>0.50</td>
<td>BPL</td>
<td>370</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>DES</td>
<td>492</td>
<td>60</td>
</tr>
<tr>
<td>0.75</td>
<td>BPL</td>
<td>494</td>
<td>367</td>
</tr>
<tr>
<td></td>
<td>DES</td>
<td>738</td>
<td>443</td>
</tr>
<tr>
<td>0.90</td>
<td>BPL</td>
<td>494</td>
<td>367</td>
</tr>
<tr>
<td></td>
<td>DES</td>
<td>738</td>
<td>443</td>
</tr>
</tbody>
</table>

---

7See [6] for a more extensive discussion of these differences.
mechanism of Weak SLD resolution. Our implementation performs this task by compiling FuzzyDES programs into core Datalog in order to obtain a set of Datalog rules that allows the emulation of the mentioned Weak SLD resolution procedure. We have precisely defined two processes for generating such compiled programs, leading to what we call the BPL expanded program and the FDES expanded program. By using program transformation techniques, we have studied the relationships between both expansions. Also we prove the semantic equivalence between the WSLD rule, applied to a logic program II, and the operational mechanism of Definition 5.1, when it is applied to a BPL or FDES expanded program II' obtained from II. Moreover, we performed an experimental analysis of the system for performance and scalability, comparing the proposed expansions. We have provided a working system able to solve queries and commands on fuzzy deductive databases and shown some of its features with an application example. Since the DES system has been used already for big data applications (deductive data warehouses and OLAP [36] and their performance [37]), we expect its fuzzy extension also amenable for such applications. To the best of our knowledge, this is the only publicly available system implementing an interactive fuzzy deductive database with user-defined fuzzy relations.

REFERENCES