Three Decades of Research
on Database Design at PUC-Rio

Marco A. Casanova, Simone D. J. Barbosa,

Karin K. Breitman, Antonio L. Furtado

Departamento de Informática

Three Decades of Research on Database Design at PUC-Rio

Marco A. Casanova, Simone D. J. Barbosa, Karin K. Breitman, Antonio L. Furtado

{casanova,simone,karin,furtado}@inf.puc-rio.br

Abstract: Research on database design at PUC-Rio dates back to the late seventies and covers a broad range of topics, from the early development of the relational model to recent applications of semiotic concepts to the design and specification of information systems. This paper briefly reviews some of the major contributions of the group, from the perspective of the authors. It organizes the contributions according to the data model or to the underlying disciplines that they are based on. Within each section, the presentation follows a chronological order as much as possible.

Keywords: Database Design, Relational Model, Entity-Relationship Model, RDF, Description Logics, Semiotics.

Resumo: A pesquisa sobre projeto de bancos de dados na PUC-Rio data do final dos anos setenta e cobre uma vasta gama de tópicos, desde o início do desenvolvimento do modelo relacional até aplicações recentes de conceitos de semiótica ao projeto e especificação de sistemas de informação. O presente artigo passa em revista algumas das principais contribuições do grupo, a partir da perspectiva dos autores. As contribuições estão organizadas de acordo com o modelo de dados ou com as disciplinas em que se baseiam. Dentro de cada seção, a apresentação segue, na medida do possível, uma ordem cronológica.

Palavras-chave: Bancos de Dados, Modelo Relacional, Modelo Entidades-Relacionamentos, RDF, Lógica de Descrição, Semiótica.

1. INTRODUCTION

Database design constitutes, since the early seventies, a key research area at the Department of Informatics of PUC-Rio. Initially focused on relational model theory and practice, the scope of our work was soon enlarged to encompass the static, dynamic and behavioral conceptual level design of information systems with a database component, and, by introducing a plan-recognition / plan-generation paradigm, the development of methods and tools for the production of executable specifications.

Several complementary formalisms were investigated aiming at the preservation of integrity constraints. Later efforts were directed to schema-integration problems, query methods conducive to cooperative responses, and to issues arising with the transition from the closed world of database tradition to the open world of the Web. Our recent work turned to the pragmatic aspects of information systems, and exploits, for this purpose, some fundamental semiotic concepts.

The paper briefly reviews some of the major contributions of the group, from the perspective of the authors. It organizes the contributions according to the data model or to the underlying disciplines that they are based on. Within each section, the presentation follows a chronological order as much as possible.

Section 2 contains the major contributions directly related to the relational model. Section 3 covers research about the entity-relationship model. Section 4 collects contributions that are based on several formalisms, including algebraic specification. Section 5 comprises work on the cooperative behavior of information systems. Section 6 addresses applications of plan generation and recognition to the design of database systems. Section 7 focuses on research motivated by the problem of publishing databases on the Web, which is related to the use of RDF and Description Logics. Section 8 discusses a semiotic approach to the conceptual specification of information systems. Finally, Section 9 contains concluding remarks.

2. RELATIONAL MODEL

Contributions to the relational model dates back to the 1977 SIGMOD Conference, where an algebra of quotient relations was proposed [Furtado and Kerschberg 1977]. The algebra operates on partitioned relations, achieved by defining equivalence relations on n-ary relations, and is as powerful as the original relational algebra, but has the advantage of a set-processing capability.

Also in the context of data manipulation, a variant of Dynamic Logic was introduced to formalize a relational database description language (DDL) and a data manipulation language (DML) [Casanova and Bernstein 1980]. The DDL is a many-sorted first-order language that accounts for data aggregations, which were largely ignored in formalizations due to the technical difficulties of dealing with aggregation in first-order logic. The DML features a many-sorted assignment in place of the usual data manipulation statements, in addition to the normal programming language constructs.

Early research on database theory concentrated on normalization and the theory of data dependency, such as functional dependencies and multivalued dependencies, as described in database textbooks. However, such research was based on the so-called *universal relation assumption*, viz., that all relations are projections of a single *universal relation*. As a reaction to this rather unrealistic assumption, the group started investigating the design of relational databases based on horizontal decompositions, in one direction, and on functional and inclusion dependencies (which can say, for example, that every manager is an employee), in another direction.

The design of relational databases based on horizontal decompositions was addressed in [Furtado 1981]. The strategy adopted relation partitioning, already explored in the algebra of quotient relations.

The design of relational databases based on functional dependencies (FDs) and inclusion dependencies (INDs) was addressed in [Casanova *et al*. 1984a]. A simple complete axiomatization for INDs was presented, and the decision problem for INDs was shown to be PSPACE-complete. It was shown that finite implication (implication over databases with a finite number of tuples) is the same as unrestricted implications for INDs, although finite implication and unrestricted implication are distinct for FDs and INDs taken together. This was one of the first results of this kind, which called attention to a new database theory area that focused on finite implication. Furthermore, it was shown that, although there are simple complete axiomatizations for FDs alone and for INDs alone, there is no complete axiomatization for FDs and INDs taken together, in which every rule is k-ary for some fixed k (and in particular, there is no finite complete axiomatization). This was an unusual result in the sense that apparently a simple class of dependencies (FDs and INDs) can exhibit very complex and unexpected interactions.

Motivated by the above formal analysis, an investigation was undertaken on how to efficiently enforce inclusion dependencies and referential integrity, as reported at the 14th and 15th VLDB Conferences [Casanova *et al.* 1988; 1989]. An analysis of the integrity constraints defined in the SQL ISO standard in the light of the entity-relationship model was also carried out [Laender *et al*. 1994]. The results pointed out what features of integrity constraints in SQL support which features of the entity-relationship model.

Departing from the tradition of data dependencies, a database description framework [Castilho *et al.* 1982; Casanova and Furtado 1982] was introduced that accounts for both static constraints, that is, constraints on what data can be stored, and transition constraints, that is, constraints on how data can be updated. Two levels of specification were considered. At the first level of specification, a database description does not indicate how the database will be updated, and transition constraints are specified with the help of a variant of Temporal Logic. By contrast, at the second level of specification, a database description includes a set of built-in update operations, which are described by their properties. The advantages accrued from this approach are twofold: first-level specifications give a stable description of constraints, while second-level specifications suggest a strategy to enforce constraints.

The view update and the view integration problems were addressed in several papers. The view update problem refers to the question of determining how update operations expressed in terms of the views affect the underlying database. The effects of a wide range of update operations on relational views were investigated [Furtado *et al*. 1979] to identify which operations must be prohibited in order to assure harmonious interactions among database users, and which operations could be allowed, even though the structure of the view may differ substantially from the actual structure of the database.

Later on, a survey on the view update problem was published [Furtado and Casanova 1985], covering the two basic approaches proposed at that time to solve the problem. The first approach suggested treating views as abstract datatypes so that the definition of the view includes all permissible view updates, together with their translations. The second approach led to general view update translators and was based either on an analysis of the conceptual schema dependencies or on the concept of view complement to disambiguate view update translations. The first approach anticipated much of the ideas underlying the persistency toolkits now available.

View integration was investigated with the help of three classes of interrelational dependencies, inclusion dependencies, exclusion dependencies and union functional dependencies [Vidal and Casanova 1983]. The process of view integration was divided into two steps, combination and optimization. View combination consists in defining new interrelational dependencies that capture similarities between different views. The optimization step tries to reduce redundancy and the size of the schema. General results about interrelational dependencies were also developed, leading to an optimization procedure for a restricted class of schemas.

3. ENTITY-RELATIONSHIP MODEL

Results about the entity-relationship model were already reported at the 1st ER Conference [Santos *et al.* 1979]. A datatype approach for database semantics was considered using the ER model as a framework. The approach introduced three datatype constructors to derive complex datatypes in the model: sum, product and correspondence. It was shown how these constructors could be used to produce semantically relevant abstractions, which impose rules on the behavior of the database. Two kinds of such rules were distinguished: existence constraints and operational constraints. The interactions among these two kinds of constraints were then investigated.

At the 2nd ER Conference, a method, also based on abstract datatypes, was proposed for representing a database application on a simple entity-relationship data model [Furtado *et al.* 1981]. The method included the verification and testing of the representation, which was simplified by the usage of procedural specifications.

At the 8th ER Conference, two constructs that capture and extend the generalization and subset abstractions were proposed [Tucherman *et al.* 1989], together with operations to maintain entity and relationship sets organized according to these constructs. The semantics of the constructs and operations were carefully defined, bringing up some key aspects of the entity-relationship model.

As a result of the investigation on the ER model, an expert software tool, called CHRIS, was developed to help in the design and rapid prototyping of information systems containing a database component [Tucherman *et al.* 1990]. The tool translated an ER conceptual specification, produced in the course of a dialogue with the designer, into relational tables, which were then created via an interface with a relational DBMS. The tool also included a module which enforced the integrity constraints of the application.

At the 9th ER Conference, two papers were presented. The first paper [Furtado and Casanova 1990] described a declarative way of specifying both the structure and the operations of an entity-relationship schema. The specification was in fact executable. The paper proceeded to describe a plan generation algorithm and a method to introduce the time dimension, whereby the facts that hold at a certain instant can be inferred from the record of the operations executed. By combining these features, the paper showed how to extend temporal databases so as to cover past, present and future states (as determined by the fixed commitments), as well as to draw plans coupled with time schedules.

The second paper [Casanova *et al.* 1990] defined a design algorithm that accepts as input an entity-relationship conceptual schema and generates an optimized relational representation for the schema (optimized in the sense that the number of dependencies of the relational schema is minimized). The paper proceeded to describe a redesign algorithm that accepts as input a conceptual schema, the relational representation for the schema resulting from the design algorithm and a sequence of changes on the schema, and produces as output the new conceptual schema and a plan to create an optimized relational representation for the new schema and to restructure the database state accordingly. An expanded version of the paper was published in DKE later on [Casanova *et al.* 1993].

The question of database redesign was retaken at the 15th ER Conference [Silva *et al.* 1996]. A new approach for restructuring databases was introduced, based on the generation of a transient virtual database state that is used to construct the new database state and that can be obtained without modifying the current relational representation. The method provided means to collect additional data and to integrate such data to the current database state, as well as to check if all the new integrity constraints are satisfied.

The mapping strategy outlined in [Laender *et al.* 1994] was generalized in [Silva *et al.* 2000]. The new strategy introduced key pairing constraints, which must be enforced to guarantee the correctness of the mapping. The paper also characterized a class of ER schemas for which this generalized strategy produces relational schemas that correctly capture the semantics of specialization structures.

Finally, a recent survey, included in the Encyclopedia of Database Systems, summarizes work on mapping entity-relationship schemas into the relational schemas [Borgida *et al.* 2009].

Further contributions to the ER conferences are reported elsewhere in this paper. This long tradition of contributions to the ER Conferences was kindly recognized by the organizers of the 28th ER Conference through an invited talk [Furtado *et al.* 2009].

4. FORMAL SPECIFICATION AND MODULARIZATION

The earliest contribution to database design based on algebraic specifications was published in 1981 [Santos *et al.* 1981]. The paper proposed a formalism adequate for the specification of behavioral properties of data bases. Both update and query requests were modeled in the language of the formal system, and were uniformly treated as a theorem proving process.

We may also include as a contribution to formal database specification a practitioner-oriented paper first published in the ACM/SIGMOD Record [Furtado 1983; Furtado and Maibaum 1985]. The paper helped bridging the gap between theoretical research and the real practice of database specification. The paper argued that, during the database design process, three questions must be answered: (a) What information will be needed? (b) How is the information going to be used? (c) Under what form will the information be kept? The aspects underlying these questions are, respectively, information, operations and representation. The answers should be satisfactory, in the sense that the information must be meaningful, valid and useful to the activities of the institution; the information should also be easily and efficiently accessible.

Research then proceeded in three directions: complementary specifications, stepwise refinement and modular design.

At the 7th VLDB Conference, a methodology was proposed for the systematic derivation of a series of complementary specifications of a database application [Veloso *et al.* 1981]. The starting point for the series was chosen so as to be obtainable without undue difficulty from an informal specification. Thereafter, each formal specification was systematically derived from the preceding one. Having distinct aims, the specifications jointly provide a multi-purpose, comprehensive characterization of the data base application.

The topic of complementary specifications was retaken in a paper presented at the 3rd PODS Conference [Casanova *et al.* 1984b]. The major contribution of the paper lied in selecting the appropriate variation of each formalism for each level of specification, in the style of organizing the formalisms together into a coherent conceptual design framework, and in the formal notion of refinement binding the different levels. Thus, contrarily to most published literature, the paper neither limited itself to just one formalism, at just one level, nor forced the use of the same formalism at different levels, which often creates distortions.

Logical, algebraic, programming language, grammatical and denotational formalisms were investigated with respect to their applicability to formal database specification in a closely related paper [Veloso and Furtado 1985]. On applying each formalism for the purpose that originally motivated its proposal, the paper showed that they all have a fundamental and well integrated role to play in different parts of the specification process.

Stepwise refinement and modularization was addressed, for example, in [Schiel *et al.* 1984]. Modularization was discussed as another dimension in the specification process, orthogonal to stepwise refinement.

Under the topic of modularization, we include a 1991 TODS paper [Casanova *et al.* 1991]. The modularization discipline incorporated both a strategy for enforcing integrity constraints and a tactic for organizing large sets of database structures, integrity constraints, and operations. A software tool that helps the development and maintenance of database schemas modularized according to the discipline was also developed. The tool incorporated, in a declarative style, a description of the design

and redesign rules behind the modularization discipline, hence facilitating the incremental addition of new expertise about database design.

5. COOPERATIVE BEHAVIOR

This section covers two closely related problems: how to achieve cooperative behavior and how to avoid misconstruals.

Briefly, an information system exhibits cooperative behavior to the extent that it interacts with users in ways that: contribute to the achievement of the users' goals and plans; keep the users' understanding of the system in harmony with the definition and contents of the system; conform to the established integrity and authorization constraints.

In particular, when interacting with a database, a user may be tempted to infer further information from that explicitly obtained from previous queries. However, since his world model is often faulty or incomplete, a fact he infers may be false with respect to the database. Such facts are often called *misconstruals*. For example, a client of a DVD rental store, after consulting the catalog and verifying that a certain DVD has not been rented, may inadvertently infer that the DVD is available, when it has actually been reserved. A more cooperative system would reply that the DVD is in store, but not available, since it has been reserved.

At the 10th ER Conference, in order to achieve cooperative behavior, an algorithm was proposed that does not execute a request literally, but rather transforms the request appropriately, guided by a set of modification rules [Hemerly *et al.* 1991a]. The algorithm may modify a request by invoking rules before the request is actually executed, after the request is successfully executed, or even after a failed execution. As part of project NICE, a prototype tool was implemented to run experiments with the request modification algorithm, which incorporated a plan and schedule generation algorithm, a temporal database package, a query-the-user facility and a session-monitoring feature [Casanova and Furtado 1990].

These ideas were expanded later on into a model of question-answering [Perez Alcazar *et al.* 1994; Sena and Furtado 1998]. Taking as the starting point an input query, the system answers the query and then, in the course of a dialogue, tries to suggest new queries related to the input query. The dialogue control was based on the structure of the concepts stored in the knowledge base, on domain restrictions, and on specific constraining rules.

The problem of avoiding misconstruals was addressed in two papers that follow closely related strategies, but which differ on the formalisms used. The strategy was to create user models that capture the intuition that, whenever the user needs to derive a positive fact F in his inference, he must check whether the current log does not indicate that F must be rejected. The first paper [Hemerly *et al.* 1991b] described a model for users' inferences that directly checked if F must be rejected. The second paper adopted a more elegant solution based on Default Logic [Hemerly *et al.* 1993]. Both papers considered a cooperative interface which was responsible for all inferences from the deductive database necessary to answer the users' queries and to compute what additional information to include in the log to avoid misconstruals.

6. PLAN GENERATION / PLAN RECOGNITION

We have been working with the conceptual modeling of information systems with a database component, considering their static, dynamic and behavioral aspects. The three aspects are integrated through the application of a plan-recognition / plan-generation paradigm [Furtado and Ciarlini 1997; Furtado and Ciarlini 2000b].

The static aspect concerns what facts hold at some database state, conveniently described in terms of the entity-relationship model.

The dynamic aspect corresponds to events that can produce state transitions. Events result from the execution of operations, defined in a declarative style by their pre-conditions and post-conditions, according to the STRIPS proposal [Fikes and Nilsson 1971]. Pre-conditions involve the presence or absence of facts, whereas post-conditions comprise the sets of facts added or deleted as the effect of the operation. Adopting the notion of abstract data types, implicit in object-oriented approaches, we require that facts can only be modified through the execution of such operations, whose pre- and post-conditions are adjusted so as to preserve all integrity constraints.

The behavioral aspect refers to the agents authorized to cause events by performing the operations. To model this aspect we mainly use goal-inference rules, which indicate what facts should hold, or cease to hold, at a target state that an agent will be motivated to bring about in view of a situation, again expressed in terms of facts holding or not holding, prevailing at the current state [Ciarlini *et al.* 2000]. In order to reach the desired target state, an agent would execute − or ask the authorized agents to execute − some appropriate plan, composed of one or more pre-defined operations. As a further development, we have started to look at agent profiles involving three kinds of personality factors, from which a decision-making process could operate: drives for the emergence of goals from situations, attitudes for the choice of plans to achieve the preferred goal, and emotions to decide whether or not to commit to the execution of the chosen plan, depending on the expected emotional gain when passing from the current to the target state [Furtado and Ciarlini 2002; Barbosa *et al.* 2010]. And, as an inducement to revise individual decisions, we included competition and collaboration interferences, as prescribed for multi-agent contexts [Willensky 1983].

In order to make our conceptual specifications executable, we created an environment where entity and relationship classes, operations, and goal-inference rules and agent profiles are all represented as Prolog clauses. Also written in Prolog, algorithms are provided for planning and for the simulated execution of the generated plans [Ciarlini and Furtado 1999a,b; Ciarlini and Furtado 2002]. Moreover it was noted that simulation can become a useful resource to support learning or training [Ciarlini and Furtado 2003].

The plan-recognition side of the paradigm is relevant, after the system has been made operational, as a means to extend conventional query facilities towards truly cooperative responses. Cooperation, as discussed in section 5, is most effective when one can detect what the user is trying to accomplish. The plan-recognition algorithm, which we adapted from [Kautz 1991], matches a few observed actions of the user against a library of previously recorded typical plans. The observed actions can be taken from the execution log, which is updated whenever each operation of a transaction of the user's initiative is executed. As we explained in [Furtado and Ciarlini 2001], the library of typical plans, in turn, can be constructed by inspecting the log and extracting and filtering sequences of executed operations whereby the transition indicated in some goal-inference rule has been achieved.

Treating databases as a component of information systems encompassing facts, events and agents permits a shift from a purely descriptive to a narrative context [Furtado 1999]. Indeed in a workshop devoted to the application of natural languages to information systems, we showed how to generate template-based natural language text, by inspecting the plot-structured execution log and analyzing it against our three-level conceptual schemas [Furtado and Ciarlini 2000a]. It is therefore not surprising that all the discussion in this section applies in essentially the same way to literary genres [Furtado and Ciarlini 1999; Furtado 2004; Ciarlini *et al.* 2010], whenever the fictional events can equally be attributed to a pre-defined repertoire of operations performed by agents (cf. the functions and *dramatis personae* in [Propp 1968]). Recognizing that literary genres ruled by identifiable conventions can thus be treated as one more kind of application domain, we have adopted plan-based plot composition, coupled with several dramatization techniques and visual media, within an ongoing digital storytelling project [Ciarlini *et al.* 2002; Ciarlini *et al.* 2005].

7. PUBLISHING DATA ON THE WEB

This section covers recent research motivated by the problem of designing databases to be exposed on the Web. It first covers schema matching, which is part of the problems one has to face when designing mediated schemas. Then, it discusses the design of the constraints of the mediated schema. Finally, it briefly touches on the problem of accessing and publishing databases on the Deep Web. Some of the ideas covered in this section were presented at the XXII Brazilian Symposium on Data Bases as an invited talk [Casanova 2007].

A recent paper [Casanova *et al.* 2007] surveys and expresses our position with respect to the most frequent approaches to address the schema matching problem, that is, the problem of defining concepts of a target schema T in term of the concepts of a source schema S.

The *syntactic approach* uses syntactical hints to match S and T. This approach depends on the implicit assumption that syntactical proximity implies semantic similarity. However, it is not difficult to convince oneself that such assumption is often unwarranted and may lead to incorrect mappings. The *semantic approach* uses semantic clues to generate hypotheses about schema matching. As a special case, a convenient approach, sometimes called *extensional* or *instance-based*, is to detect how the same real-world objects are represented in the source and target databases and to use the information thus obtained to match the schemas. This approach is grounded on the interpretation that "terms have the same extension when true of the same things", advocated by W. V. Quine in one of his most influential papers [Quine 1968].

Both the syntactic and the semantic approaches work *a posteriori*, in the sense that they start with pre-existing databases and try to match their schemas. By contrast, one may adopt an *a priori approach*, which emphasizes that, whenever specifying databases that will interact with each other, the designer should start by selecting an appropriate standard, if one exists, to guide the design of the (exported) schemas. If both the source schema S and the target schema T follow the same common schema, then matching S and T becomes trivial. Ontologies, in the sense used by the World Web Consortium - W3C, come in hand to transform the a priori approach into a viable strategy. The technology is readily available. It is just a matter of disciplined schema design.

In fact, we observe that disciplined schema design for the Web is in fact one of the basic ideas behind the Linked Data Initiative, and it should be one of the design guidelines of triplification tools (that is, tools that publish databases as sets of RDF triples), as exemplified in [Salas *et al.* 2010].

The problem of matching two schemas that belong to an OWL dialect is addressed in [Leme *et al.* 2009; Leme *et al.* 2010], following an instance-based approach. The schema matching problem is decomposed into the problems of *vocabulary matching* and *concept mapping*. The paper also introduced sufficient conditions guaranteeing that a vocabulary matching induces correct concept mappings. Research along these lines was also reported, for example, in [Brauner *et al.* 2008; Leme *et al.* 2008; Gomes *et al.* 2010].

We now turn to the design of the constraints of the mediated schema. Recall that a mediation environment contains export schemas, a mediated schema and mappings between them. The constraints of the mediated schema are relevant for a correct understanding of what the semantics of the external schemas have in common.

The problem of changing the constraints of the mediated schema is formulated in [Casanova *et al.* 2010] as a problem of computing the greatest lower bound (g.l.b.) of the theories induced by two sets of constraints (the g.l.b. of two theories is defined as their intersection). Then, for an expressive family of conceptual schemas, formalized in Description Logic, the paper showed how to efficiently decide logical implication and how to compute a representation of the g.l.b. of two theories induced by two sets of constraints. The decision procedure essentially explores the structure of a set of constraints, captured as a graph. The procedure to compute the greatest lower bound of two theories induced by two sets of constraints is a direct consequence of the decision procedure. These results are extended in [Casanova *et al.* 2011] to more expressive families of schemas, also formalized in

Description Logic. Interestingly, the limitations here are reminiscent of the problems created by the interaction of functional and inclusion dependencies, reported in [Casanova *et al.* 1984a].

Finally, we briefly touch on the problem of accessing and publishing databases on the Deep Web. Recall that, unlike the Surface Web of static pages, the Deep Web comprises data stored in databases, dynamic pages, scripted pages and multimedia data, among other types of objects. Deep Web databases are typically under-represented in search engines due to the technical challenges of locating, accessing, and indexing the databases.

Two basic approaches to access Deep Web data have been proposed. The first approach, called surfacing or Deep Web Crawl, tries to automatically fill out HTML forms to query the databases. Queries are executed offline and the results are translated to static Web pages, which are then indexed. The second approach, called federated search or virtual integration, suggests using domain-specific mediators to facilitate access to the databases. Hybrid strategies, which extend the previous approaches, have also been proposed.

A different approach to publish Deep Web databases was proposed in [Piccinini *et al.* 2010]. The basic strategy consists of creating a set of natural language sentences, with a simple structure, to describe Deep Web data, and publishing the sentences as static Web pages, which are then indexed as usual. The use of natural language sentences is convenient for three reasons. First, they lead to Web pages that are acceptable to Web crawlers that consider words randomly distributed in a page as an attempt to manipulate page rank. Second, they facilitate the task of more sophisticated engines that support semantic search based on natural language features. Lastly, the descriptions thus generated are minimally acceptable to human users.

8. SEMIOTICS

8.1 Overview

A semiotic view of the conceptual specification of information systems recognizes that language, and, more generally, any communication process goes through four levels: lexical, syntactic, semantic, and pragmatic [Morris 1938/1970]. At the semantic level, the correspondence between the stored data and real world facts is considered, but to design systems of practical usefulness, one still needs to investigate what purposes they will serve, which falls in the scope of the pragmatic level.

Accordingly, our conceptual design method encompasses not only facts, but also events and agents. Motivated by their goals, defined in terms of database facts, agents try to cause the occurrence of events whereby a database state is reached where the goals are satisfied. As already reported in Section 6, our plan-recognition / plan-generation paradigm covers all these aspects and leads to executable specifications, which allow simulation experiments to effectively test the proposed conceptual design.

Next, we studied how the information space is structured, and again notions developed in the area of semiotics proved to be helpful. Based on Burke's four master tropes proposal [Burke 1969], we were able to identify four modalities of semiotic relations, which provide an understanding – with several applications – of the kinds of connections that may exist not only between facts, but also between events and between agents.

Among these semiotic relations, the paradigmatic relation deserved special attention in the course of our project. Paradigms and the associated notions of similarity and analogy constitute the basis for handling classes, as well as best-match search and classification [Breitman *et al.* 2007b]. Analogy, in particular, creates opportunities to handle information across various domains, as Web environments and Internet search-engines especially favor.

8.2 Semiotic Relations

Based on Burke's discussion [Burke 1969] of what he calls the four master tropes – metonymy, metaphor, irony and synecdoche –, we identified four types of relations that can exist not only between facts, but also between events and between agents – syntagmatic, paradigmatic, antithetic and meronymic relations, respectively (Figure 1) [Ciarlini *et al.* 2008; Furtado *et al.* 2009].

Syntagmatic relations express contiguity or coherence, and justify putting elements in a sequence. For example, it makes sense to align the various properties of an entity instance when describing it, or to mention together entity instances between which some relationship holds. And if the post-conditions of an operation satisfy pre-conditions of another operation, the two operations can appear in the same plan, in that order.

Paradigmatic relations express similarity, thus justifying placing elements in the same sets, as also to replace one element by another in the same position of a sequence. For example, entity classes are formed of instances having a number of properties in common. And alternative plans can be obtained by replacing some operation by another that can also be executed at that stage and produce the intended effect.

Antithetic relations express negation or opposition. Informally speaking, syntagms and paradigms constitute axes in a two-dimensional space, whose bounds are imposed by incompatibilities preventing certain combinations of elements. Integrity constraints is the first example that comes to mind, e.g., forbidding the participation in a relationship of an entity instance with a value in a specified range for a given attribute. Also, an operation in a plan can produce effects that conflict with the pre-conditions of an operation that is being considered for inclusion in a subsequent position.

Meronymic relations have to do with level of detail, and are crucial for modularization. The semantic is-a and part-of hierarchies between entity classes provide a familiar example. Already at the first conference wherein the ER model was discussed [Santos 1979], we demonstrated how the model could be extended with these hierarchies, which in fact induce a partial order lattice-like structure [Barbosa *et al.* 2007], possibly with multiple inheritance of properties, whenever classes with upward links with more than one parent class are present (e.g. trainees as a specialization of both students and employees). Plans can also be represented in this double hierarchy (or lattice-like structure, as in [Kautz 1991] and in our own work [Furtado and Ciarlini 2000b]), with single operations as the smallest components.
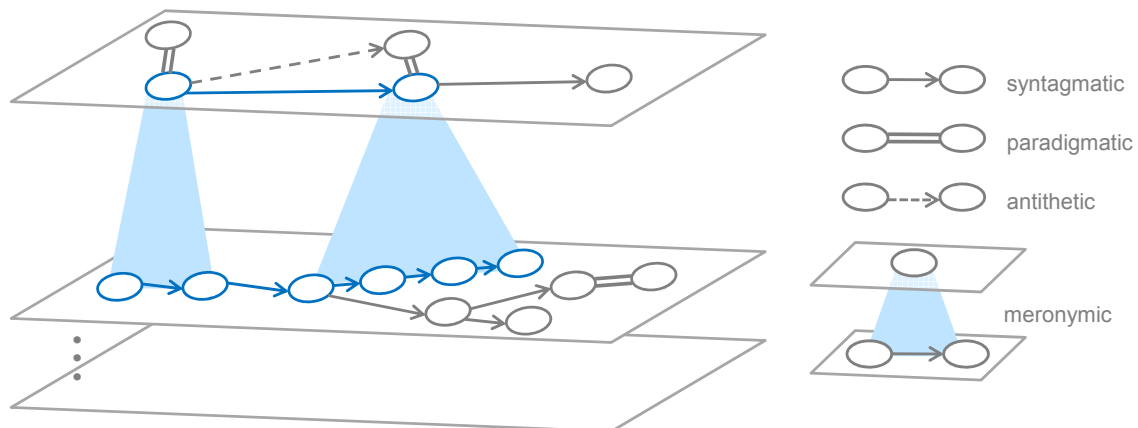

Figure 1: The four semiotic relations in an information space

Notice that, if two elements are specializations of the same more general element in an is-a hierarchy, there is a paradigmatic relation between the two specialized elements [Furtado 1992]. On the other hand, part-of hierarchies presuppose syntagmatic relations between the elements into which an element is decomposed. Burke himself had noted that the four master tropes are not mutually independent and therefore it is not surprising that the four relations derived from the tropes interact in

significant ways, as we have already noted with respect to the negative interference of the antithetic relations.

The four semiotic relations are also applicable to agents. A useful characterization can be done on the basis of an analysis of goal and plan interferences, as in [Willensky 1983], and of the list of *dramatis personae* of [Propp 1968]. Agents that collaborate with each other (Propp's helpers and donators) are in syntagmatic relation, agents who play the same role or can be mutually replaced are in paradigmatic relation, competing agents or agents whose plans aim at opposite results (villains, false heroes) are in antithetic relation, teams or organizations with subdivisions are in meronymic relation.

To organize data from different sources, a requirement made more pressing by the advent of the Web, we preferred to consider frames, a data structure of ample use in artificial intelligence applications [Barbosa *et al.* 2007]. Notice that aligning the known property-value pairs of an entity instance in the same frame reflects their belonging to the same syntagm. On the other hand, sets of frames are justified if the frames are associated by some paradigm. Well in agreement with our pragmatic orientation, sets of frames that are apparently very heterogeneous can make good sense: for example, consider frames with air tickets information kept together with frames for hotel reservations, having in common their pertinence to an *ad-hoc* paradigm, such as their being associated with the same excursion tour.

We use a data structure, called plot, for representing plans [Furtado *et al.* 2008]. Plots, which are also frame-like, register the component operations, each with a tag, and, in addition, a list of tag-pairs to express the partial order imposed by the specified pre-conditions. As happens with frames, plots and sets of plots are formed, respectively, in view of syntagmatic and paradigmatic relations.

We have defined and have experimented with a frame algebra and a plot algebra [Furtado *et al.* 2009; Karlsson *et al.* 2009], both with a repertoire of seven operators that is complete in the sense that the operators allow all modifications justified by the four semiotic relations: for the syntagmatic relation, product and projection; for the paradigmatic relation, union and selection; for the antithetic relation, difference; for the meronymic relation, combination and factoring. This can be compared, with the changes required by the different data structures, to classic relational algebra, except that the meronymic relation can only be handled in NF2 (non-first-normal-form) tables, an early proposal in this direction being our work (mentioned at the beginning of section 2) on quotient relations [Furtado and Kerschberg 1977].

8.3 Similarity and Analogy

We remarked in the previous section that paradigmatic relations justify the placement of entity instances in the same entity class. The usual paradigm is the presence of the same properties, i.e. attributes and relationships, in all instances. For some properties it may be allowed that the value be currently unknown or, in some cases, inapplicable. For other properties, default values may be assumed, or a specific value or value range may be prescribed.

Now, when searching for an instance having certain indicated properties, one may fail to find an exact match. We decided to extend our query facilities by searching instead for the best possible match, which may fail to be exact but should lie within a minimum threshold in terms of similarity. For example, one may define a similarity measure based on the number of required properties that are present regardless of their values, number of coincident values for the properties whose values are explicitly indicated, and number of value conflicts. Weights can be applied when computing the result.

A similarity measure criterion was also used for, given an entity instance whose class is ignored, and for which some properties or property-values are also not known, finding through a breadth-first strategy to which class it best fits in [Barbosa *et al.* 2007]. Our classification algorithm works on a lattice-like structure of classes, starting from a top "universal class" and going down in an attempt to find a best-matching class. Notice that, since is-a links impose the inheritance of properties, the algorithm will not go further down once a class is found having all specified properties – obviously

the classes below would have such properties, but would also have others whose presence in the instance under examination should not be unwarrantedly assumed.

The notion of similarity and similarity measures is fundamental for the determination of the classes themselves. One may start by looking at sets of unclassified instances, assign names (preferably taken from a standard vocabulary) to their properties, and then apply a clustering algorithm, using similarity measures as clustering criterion [Nunes and Casanova 2010]. Then classes can be chosen by identifying clusters with high internal cohesion and low external connectivity. The algorithm will also identify centroids, which would then be selected as prototypical elements. This is in agreement with Lakoff's thesis [Lakoff and Johnson 1980; Lakoff 1987], which claims that human beings classify entity instances verifying how close they are to what in the person's cultural environment is considered a typical element of the class.

For explanatory purposes, we chose to use the term analogy to denote similarity between instances of different domains. Analogy may exist even when totally different objects are involved. A classic example is the metaphor whereby organizations are said to be like machines. In fact, there is a structural coincidence: machines have pieces, organizations have employees. The coincidence is also functional: if pieces have defects they can be either repaired or replaced. If employees are not working properly, they can be trained or replaced. The example shows how analogy can be detected and then utilized for a practical end. First we identify a mapping between the elements; in this case employees correspond to pieces, the property of not working well corresponds to being defective. Then, as a consequence, we can look for suitable plans to change the undesired property by adapting the corresponding plans supposedly already known for the other domain, which incidentally is reminiscent of the case-based reasoning strategy [Kolodner and Leake 1996] investigated in the artificial intelligence area.

Indeed, we have proposed a method to help designing a target class schema using as a source some typical well-known class, possibly in a distinct domain, whose schema exhibits an analogous structure [Breitman *et al.* 2007a; Breitman *et al.* 2007b; Casanova *et al.* 2008; Furtado *et al.* 2008]. Since, besides whatever the source and the target classes have in common, several kinds of conflicts are also likely to occur, some adaptation heuristics is required, a very promising semiotic strategy that we have studied being referred to as blending [Fauconnier and Turner 1994]. For example, the somewhat complex case of weak entities has as canonical example the case of the dependents of an employee, which can be used, after suitable mappings and adaptations, to design the schemas needed for another case of weak entity. Once this is done, our algorithms are also able to derive plans by locating and adapting plans kept in the plan library developed for the source schemas.

Finally, it should be mentioned, always in accordance with our pragmatic orientation, that similarity or analogy has a much wider scope than one might imagine at first. For example, entity instances are comparable not only in view of their properties but also in view of the executed plans in which they were involved. Thus, employees may be compared in terms of career patterns.

9. CONCLUDING REMARKS

This paper spans over thirty years of research on database design and covers a broad range of topics. Still it does not represent, due to the usual space limitations, a thorough summary of all contributions of the group to database research and to other related areas. The last section discusses recent work which departs from the more familiar database research tradition. However, it brings to the front semiotic concepts that help attribute meaning to terms such as the Pragmatic Web, for example.

Besides the research papers surveyed in the preceding sections, the authors published a few books, some of which served to transmit either the more general concepts or their personal orientation to the graduate students they helped prepare along this long period, who eventually initiated their own research groups in several leading Brazilian universities. Among such books, we choose to mention [Casanova 1981], containing the text of the author's doctoral thesis, [Furtado and Neuhold 1986],

discussing two distinct approaches to formal database design, and [Breitman, Casanova and Truszkowski 2007], surveying in depth various topics pertaining to the Semantic Web technology, such as ontologies. As textbooks for undergraduate and introductory graduate courses, we started long ago with [Furtado and Santos 1979], describing the early hierarchic, network and relational data models, as well as file organizations and the then available database management systems, and, more recently [Barbosa and Silva 2010], giving a comprehensive coverage of the Human-Computer Interaction area, with an emphasis on semiotic aspects, whose relevance to our work has been stressed in the present paper.

We are pleased to acknowledge our long standing collaboration with the database groups at the Department of Computer Science, UFMG, the Department of Computing, UFC, the Department of Applied Informatics, UNIRIO, and the Image Processing Division, INPE, to name just a few. A large part of the references listed below reflects their invaluable participation. As a further incentive, the INCT Program on Web Science is now offering a continuing opportunity for collaboration, especially on the topics related to the last two sections.

References

BARBOSA, S.D.J., BREITMAN, K.K., CASANOVA, M.A., FURTADO, A.L. Similarity and Analogy over Application Domains. In: Proc. 22nd Brazilian Symposium on Data Bases, João Pessoa, Brazil, pp. 238 – 254, 2007.

BARBOSA, S.D.J., FURTADO, A.L., CASANOVA, M.A. A decision-making process for digital storytelling. In: *IX Brazilian symposium on games and digital entertainment*, 2010.

BARBOSA, S.D.J., SILVA, B.S. *Interação Humano-Computador*. Rio de Janeiro: Campus/Elsevier, 2010.

BORGIDA, A., CASANOVA, M.A., LAENDER, A.H.F. Logical database design: from conceptual to logical schema In: Encyclopedia of Database Systems. Heidelberg: Springer-Verlag, pp. 1645 – 1649, 2009.

BRAUNER, D.F., GAZOLA, A., CASANOVA, M.A., BREITMAN, K.K. Adaptative Matching of Database Web Services Export Schemas. In: Proc. 10th International Conference on Enterprise Information Systems, Barcelona, pp.49 – 56, 2008.

BREITMAN, K.K., BARBOSA, S.D.J., CASANOVA, M.A., FURTADO, A.L. Using analogy to promote conceptual modeling reuse. In: *Workshop on leveraging applications of formal methods, verification and validation*, 2007a.

BREITMAN, K.K., BARBOSA, S.J., CASANOVA, M.A., FURTADO, A.L. Conceptual Modeling by Analogy and Metaphor In: Proc. ACM 16th Conference on Information and Knowledge Management (CIKM 2007), Lisboa, 2007b.

BREITMAN, K.K., CASANOVA, M.A., TRUSZKOWSKI, W. *Semantic Web: Concepts, Technologies and Applications*. NASA Monographs in Systems and Software Engineering. Heidelberg: Springer-Verlag, p.337, 2007 (ISBN 978-1-84628-581-3).

BURKE, K. *A Grammar of Motives*. University of California Press, 1969.

CASANOVA, M.A. *The Concurrency Control Problem for Database Systems*. Lecture Notes in Computer Science, Vol. 116. Heidelberg: Springer-Verlag, p. 175, 1981 (ISBN 3-540-10845-9).
(see http://books.google.com/books/about/The_concurrency_control_problem_for_data.html?id=I_H9oPVADKsC)

CASANOVA, M.A. Schema Matching Revised. In: Proc. 22nd Brazilian Symposium on Data Bases, João Pessoa, Brazil, 2007.

CASANOVA, M.A., BARBOSA, S.D.J., BREITMAN, K.K., FURTADO, A.L. Generalization and blending in the generation of entity-relationship schemas by analogy. In: Proc. 10th International Conference on Enterprise Information Systems, Barcelona, 2008.

CASANOVA, M.A. AND BERNSTEIN, P.A. A Formal System for Reasoning about Programs Accessing a Relational Database. ACM Transactions on Programming Languages and Systems, v.2, pp.386 – 414, 1980.

CASANOVA, M.A., BREITMAN, K.K., BRAUNER, D.F., MARINS, A. Database Conceptual Schema Matching. Computer (Long Beach), v.40, pp.102 – 104, 2007.

CASANOVA, M.A., BREITMAN, K.K., FURTADO, A.L., VIDAL, V.M.P., MACEDO, J.A.F. On the Problem of Matching Database Schemas. In: Conquering Complexity, Hinchey, M. and Coyle, L. (to appear).

CASANOVA, M.A., FAGIN, R., PAPADIMITRIU, C.H. Inclusion Dependencies and Their Interaction with Functional Dependencies. Journal of Computer & Systems Sciences International, v.28, pp. 29 – 59, 1984a.

CASANOVA, M.A. AND FURTADO, A.L. A Family of Temporal Languages for the Description of Transition Constraints. In: Proc. Workshop on Logical Bases for Databases, 1982.

CASANOVA, M.A. AND FURTADO, A.L. An Information System Environment based on Plan Generation In: Proc. Working Conference on Cooperating Knowledge based Systems, Keele, UK, 1990.

CASANOVA, M.A., FURTADO, A.L., TUCHERMAN, L. A Software Tool for Modular Database Design. ACM Transactions on Database Systems, v.16, pp.209 – 234, 1991.

CASANOVA, M.A., FURTADO, A.L., TUCHERMAN, L. Enforcing Including Dependencies and Referential Integrity. In: Proc. International Conference on Very Large Data Bases. San Francisco: Morgan Kaufmann, 1988.

CASANOVA, M.A., LAUSCHNER, T., LEME, L.A.P.P., BREITMAN, K.K., FURTADO, A.L., VIDAL, V.M.P. Revising the Constraints of Lightweight Mediated Schemas. Data & Knowledge Engineering, v.69, pp.1274 – 1301, 2010

(originally published in: Proc. 28th International Conference on Conceptual Modeling. Berlin / Heidelberg: Springer, 2009).

CASANOVA, M.A., TUCHERMAN, L., FURTADO, A.L. Optimization of Relational Schemas Containing Inclusion Dependencies In: Proc. International Conference on Very Large Data Bases, Amsterdam, pp.317 – 325, 1989.

CASANOVA, M.A., TUCHERMAN, L., LAENDER, A.H.F. Algorithms for Designing and Maintaining Optimized Relational Representations of Entity-Relationship Schemas In: Proc. 9th International Conference on Entity-Relationship Approach (ER'90), Lausanne, Switzerland, pp. 361 – 374, 1990.

CASANOVA, M.A., TUCHERMAN, L., LAENDER, A.H.F. On the Design and Maintenance of Optimized Relational Representations of Entity-Relationship Schemas. Data & Knowledge Engineering, v.11, pp. 1 – 20, 1993 (doi:10.1016/0169-023X(93)90043-O).

CASANOVA, M.A., VELOSO, P.A.S., FURTADO, A.L. Database Specification Formalisms: An Eclectic Perspective. In: Proc. 3rd ACM SIGACT-SIGMOD Symposium on Principles of Database Systems (PODS), pp. 110 – 118, 1984b.

CASTILHO, J.M.V., CASANOVA, M.A., FURTADO, A.L. A Temporal Framework for Database Specifications. In: Proc. International Conference on Very Large Data Bases, Mexico City, pp.280 – 291, 1982.

CIARLINI, A.E.M., BARBOSA, S.D.J., CASANOVA, M.A., FURTADO, A.L. Event relations in plot-based plot composition. In: *VII Brazilian symposium on computer games and digital entertainment*, 2008.

CIARLINI, A.E.M., CASANOVA, M.A., FURTADO, A.L., VELOSO, P.A.S. Modeling interactive storytelling genres as application domains. *Journal of Intelligent Information Systems*, v.35, pp. 347 – 381, 2010.

CIARLINI, A.E.M., FEIJO, B., FURTADO, A.L. An Integrated Tool for Modelling, Generating and Exhibiting Narratives. In: Proc. AI, Simulation and Planning in High Autonomy Systems, Lisboa, 2002.

CIARLINI, A.E.M. AND FURTADO, A.L. Simulating the Interaction of Database Agents. In: Proc. Database and Expert Systems Applications Conference (DEXA'99), Florence, 1999a.

CIARLINI, A.E.M. AND FURTADO, A.L. Interactive multistage simulation of goal-driven agents. *Journal of the Brazilian Computer Society*, v.2, n.6, 1999b.

CIARLINI, A.E.M. AND FURTADO, A.L. Understanding and Simulating Narratives in the Context of Information Systems. In: Proc. 21st International Conference on Conceptual Modeling, Tampere, Finland, 2002.

CIARLINI, A.E.M. AND FURTADO, A.L. Towards a Plan-based Learning Environment. In: Proc. I PGL Database Research Conference, Rio de Janeiro, 2003.

CIARLINI, A.E.M., POZZER, C.T., FURTADO, A.L., FEIJO, B. A logic-based tool for interactive generation and dramatization of stories. In: Proc. ACM-SIGCHI International Conference on Advances in Computer Entertainment Technology, Valencia, 2005.

CIARLINI, A.E.M., VELOSO, P.A.S., FURTADO, A.L. A formal framework for modelling at the behavioural level. In H. Kangassalo; H. Jaakola; E. Kawaguchi (orgs.). *Information modeling and knowledge bases XII*. IOS Press, 2000.

FAUCONNIER, G. AND TURNER, M. Conceptual projection and middle spaces. *Technical Report 9401,* University of California, San Diego, 1994.

FIKES, R. AND NILSSON, N. STRIPS: a new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, v. 2, pp. 189 – 208, 1971.

FURTADO, A.L. Narratives over real-life and fictional domains. In: Proc. 19th Brazilian Symposium on Data Bases, Brasília, 2004.

FURTADO, A.L. Horizontal Decomposition to improve a Non-BCNF Scheme. ACM/SIGMOD Record, v.12, n.1, 1981.

FURTADO, A.L. An Informal Approach to Formal Specifications. ACM/SIGMOD RECORD, pp. 45–54, 1983.

FURTADO, A.L. Analogy by generalization and the quest of the grail. ACM/SIGPLAN Notices, v.27, n.1, 1992.

FURTADO, A.L. Narratives and temporal databases: an interdisciplinary perspective. In J. Akoka; H. Kangassalo; B. Thalheim (orgs.). *Conceptual modeling: historical perspectives and future directions.* Springer, 1999.

FURTADO, A.L., BREITMAN, K.K., CASANOVA, M.A., BARBOSA, S.D.J. Applying Analogy to Schema Generation. iSYS - Revista Brasileira de Sistemas de Informação, v.1, 2008.

FURTADO, A.L. AND CASANOVA, M.A. Updating Relational Views. In: Query Processing in Database Systems. New York: Springer, 1985.

FURTADO, A.L. AND CASANOVA, M.A. Plan and Schedule Generation over Temporal Databases. In: Proc. 9th International Conference on Entity-Relationship Approach (ER 90), 8-10 October, Lausanne, Switzerland. pp. 235–248, 1990.

FURTADO, A.L., CASANOVA, M.A., BARBOSA, S.D.J., BREITMAN, K.K. Analysis and Reuse of Plots Using Similarity and Analogy In: Conceptual Modeling, Barcelona, Spain. Conceptual Modeling - ER 2008, Lecture Notes in Computer Science. Berlin / Heidelberg: Springer, v. 5231, pp.355–368, 2008.

FURTADO, A.L., CASANOVA, M.A., BREITMAN, K.K., BARBOSA, S.D.J. A Frame Manipulation Algebra for ER Logical Stage Modeling. In: Proc. 28th International Conference on Conceptual Modeling, 2009, Gramado. Berlin / Heidelberg: Springer. v. 5829, pp. 9–24, 2009.

FURTADO, A.L. AND CIARLINI, A.E.M. Plots of Narratives over Temporal Databases. In: Proc. 8th International Conference and Workshop on Database and Expert Systems Applications (DEXA'97), Toulouse, 1997.

FURTADO, A.L. AND CIARLINI, A.E.M. Operational Characterization of Genre in Literary and Real-life Domains. In: Proc. Conceptual Modeling Conference - ER'99, Paris, 1999.

FURTADO, A.L. AND CIARLINI, A.E.M. Generating Narratives from Plots using Schema Information. In: Proc. 5th International Workshop on Applications of Natural Language for Information Systems, Versailles, 2000a.

FURTADO, A.L. AND CIARLINI, A.E.M. The plan recognition / plan generation paradigm. In A. Solvberg; S. Brinkkemper; E. Lindecrona (orgs.). *Information engineering: state of the art and research themes.* Springer, 2000b.

FURTADO, A.L. AND CIARLINI, A.E.M. Constructing Libraries of Typical Plans. In: Proc. Conference on Advanced Information Systems Engineering (CAiSE), Interlaken, Switzerland, 2001.

FURTADO, A.L. AND CIARLINI, A.E.M. Cognitive and affective motivation in conceptual modelling. *Revista Colombiana de Computación*, v.3, n.2, 2002.

FURTADO, A.L. AND KERSCHBERG, L. An Algebra of Quotient Relations. In: Proc. ACM/SIGMOD Conference, 1977.

FURTADO, A.L. AND MAIBAUM, T.S.E. An Informal Approach to Formal Algebraic Specifications. Computer Journal, v. 28, 1985.

FURTADO, A.L. AND NEUHOLD, E.J. *Formal Techniques for Data Base Design.* New York: Springer-Verlag, 1986.

FURTADO, A.L. AND SANTOS, C.S. *Organização de Bancos de Dados.* São Paulo: Ed. Campus, 1979.

FURTADO, A.L., SEVCIK, K., SANTOS, C.S. Permitting Updates Through Views of Data Bases. Information Systems, v. 4, n. 4, pp. 269 – 283, 1979 (doi:10.1016/0306-4379(79)90021-8).

FURTADO, A.L., VELOSO, P.A.S., CASTILHO, J.M.V. Verification and Testing of S-ER Specifications. In: Proc. Entity- Relationship Conference, pp. 123 – 147, 1981.

GOMES, R.V., LEME, L.A.P.P., CASANOVA, M.A. MatchMaking – A Tool to Match OWL Schemas. Revista de Informática Teórica e Aplicada, v. 16, pp. 71 – 76, 2010.

HEMERLY, A. S., CASANOVA, M.A., FURTADO, A.L. Cooperative behaviour through request modification In: Proc. 10th International Conference on Entity-Relationship Approach (ER'91), 23-25 October, San Mateo, California, USA, pp. 607 – 621, 1991a.

HEMERLY, A. S., CASANOVA, M.A., FURTADO, A.L. Exploiting User Models to Avoid Misconstruals. In: Non-Standard Queries and Answers, Oxford University Press, 1995 (also in: Proc. Workshop on Nonstandard Queries and Answers, Toulouse, 1991b).

HEMERLY, A. S., CASANOVA, M.A., FURTADO, A.L. Avoiding Misconstruals in Database Systems: A Default Logic Approach. IEEE Transactions on Knowledge and Data Engineering, v. 5, Issue 6, 1993 (doi:10.1109/69.250086).

KARLSSON, B., BARBOSA, S.D.J., FURTADO, A.L., CASANOVA, M.A. A plot-manipulation algebra to support digital storytelling. In: 8th international conference on entertainment computing, 2009.

KAUTZ, H.A. A formal theory of plan recognition and its implementation. In: Reasoning about Plans. J. F. Allen et al (eds.). Morgan-Kaufmann, 1991.

KOLODNER, J.L. AND LEAKE, D. A Tutorial Introduction to Case-Based Reasoning . *In Case-Based Reasoning: Experiences, Lessons and Future Directions*, Leake, David (ed.). MIT Press. pp. 31-65, 1996.

LAENDER, A.H., CASANOVA, M.A., CARVALHO, A.P., RIDOLFI, L. An analysis of SQL integrity constraints from an entity relationship model perspective. Information Systems, v. 19, pp. 331 – 358, 1994.

LAKOFF, G. *Women, Fire, and Dangerous Things.* University of Chicago Press, 1987.

LAKOFF, G. AND JOHNSON, M. *Metaphors We Live By.* University of Chicago Press, 1980.

LEME, L.A.P.P., CASANOVA, M.A., BREITMAN, K.K., FURTADO, A.L. OWL schema matching. Journal of the Brazilian Computer Society, v.16, pp. 21 – 34, 2010.

LEME, L.A.P.P., BRAUNER, D.F., BREITMAN, K.K., CASANOVA, M.A., GAZOLA, A. Matching object catalogues. Innovations in Systems and Software Engineering, v. 4, pp. 315 – 328, 2008.

LEME, L.A.P.P., CASANOVA, M.A., BREITMAN, K.K., FURTADO, A.L. Instance-based OWL Schema Matching. In: 11th International Conference on Enterprise Information Systems, Milan, pp. 14 – 26, 2009.

MORRIS, C. W. *Foundations of the Theory of Signs.* Chicago: Chicago University Press, 1938/1970.

NUNES, B.P. AND CASANOVA, M.A. A Frame-Based System for Automatic Classification of Semi-Structured Data. Revista de Informática Teórica e Aplicada (Impresso), v. 16, pp. 87 – 92, 2010.

PICCININI, H., CASANOVA, M.A., FURTADO, A.L. W-Ray: A Strategy to Publish Deep Web Geographic Data In: 4th International Workshop on Semantic and Conceptual Issues in GIS (SeCoGIS 2010); in conjunction with the 29th International Conf. on Conceptual Modeling (ER 2010), 2010, Vancouver. Lecture Notes in Computer Science. Berlin: Springer-Verlag, v.6413. pp. 2 – 11, 2010.

PEREZ ALCAZAR, J.J., HEMERLY, A., CASANOVA, M.A., FURTADO, A.L. Cooperative interfaces for spatio-temporal databases. In: *Advances in Database and Expert Systems*, 1994.

PROPP, V. *Morphology of the Folktale.* Laurence, S. (trans.), University of Texas Press, 1968.

QUINE, W. V. Ontological relativity. The Journal of Philosophy, v. 65, n. 7, pp. 185 – 212, 1968.

SALAS, P. E., BREITMAN, K. K., VITERBO, J., CASANOVA, M. A. Interoperability by Design Using the Std-Trip Tool: an a priori approach In: Proc. 6th International Conference on Semantic Systems (I-SEMANTICS 2010), Graz, Austria, 2010.

SANTOS, C.S., NEUHOLD, E.J., FURTADO, A.L. A Data Type Approach to the Entity-Relationship Model. In: Proc. International Conference on the Entity-Relationship Approach to Systems Analysis and Design, 1979.

SANTOS, C.S., MAIBAUM, T.S.E., FURTADO, A.L. Conceptual Modeling of Data Base Operations. International Journal of Computer and Information Science, v. 10, n. 5, pp. 299 – 314, 1981 (doi: 10.1007/BF00993149)

SCHIEL, U., FURTADO, A.L., NEUHOLD, E.J., CASANOVA, M.A. Towards Multi-Level and Modular Conceptual Schema Specifications. Information Systems, v. 9, n. 1, pp. 43 – 57, 1984 (doi:10.1016/0306-4379(84)90015-2)

SENA, G.J. AND FURTADO, A.L. Correcting user queries through a cooperative database interface. In: International Conference on Computing and Information, 1993.

SENA, G. J. AND FURTADO, A.L. Towards a Cooperative Question-Answering Model. In: Proc. 3rd International Conference on Flexible Query Answering Systems (FQAS '98), London: Springer-Verlag, pp. 354 – 365, 1998.

SILVA, A.S., LAENDER, A.H.F., CASANOVA, M.A. An approach to maintaining optimized relational representations of entity-relationship schemas. In: Proc. 15th International Conference on Conceptual Modeling, Cottbus, Germany, October 7-10. Lecture Notes in Computer Science, Vol. 1157, Springer, 1996 (ISBN 3-540-61784-1).

SILVA, A.S., LAENDER, A.H.F., CASANOVA, M.A. On the relational representation of complex specialization structures. Information Systems, v. 25, n. 5, pp. 399 – 415, 2000 (doi:10.1016/S0306-4379(00)00025-9).

TUCHERMAN, L., CASANOVA, M.A., GUALANDI, P.M., BRAGA, A.P. A Proposal for Formalizing and Extending the Generalization and Subset Abstractions in the Entity-Relationship Model. In: Proc. 8th International Conference on Entity-Relationship Approach. Amsterdam: North-Holland, 1989.

TUCHERMAN, L., CASANOVA, M.A., FURTADO, A.L. The CHRIS consultant - a tool for database design and rapid prototyping. Information Systems, v. 15, Issue 2, pp. 187 – 195, 1990 (doi:10.1016/0306-4379(90)90034-M).

(also In: Proc. International Conference on the Entity-Relationship Approach, 1987).

VELOSO, P.A.S., CASTILHO, J.M.V., FURTADO, A.L. Systematic Derivation of Complementary Specifications. In: Proc. 7th International Conference on Very Large Data Bases - Volume 7, VLDB Endowment, pp. 409 – 421, 1981.

VELOSO, P.A.S. AND FURTADO, A.L. Towards Simpler and Yet Complete Formal Specifications. In: Information Systems - Theoretical and Formal Aspects (TFAIS'85), Proc. of the IFIP WG8.1 Working Conference, Sitges, Spain, 16-18 April, Elsevier/North-Holland, pp. 174 – 188, 1985 (ISBN 0-444-87706-1).

VIDAL, V.M.P. AND CASANOVA, M.A. Towards a Sound View Integration Methodology In: Proc. ACM SIGACT-SIGMOD Symposium on Principles of Database Systems, pp.36 – 47, 1983 (doi:10.1145/588058.588065).

WILLENSKY, R. *Planning and Understanding – a Computational Approach to Human Reasoning*. Addison-Wesley, 1983.