Three Decades of Research on Database Design at PUC-Rio

M.A. Casanova, S.D.J. Barbosa, K.K. Breitman, A.L. Furtado

Department of Informatics
Pontifical Catholic University of Rio de Janeiro (PUC-Rio)
Rio de Janeiro, Brazil {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.

Categories and Subject Descriptors: H. Information Systems [**H.m. Miscellaneous**]: Databases

Keywords: database design, relational model, entity-relationship model, RDF, Description Logics, semiotics

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 to database design, 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 Logic. Section 8 discusses a semiotic approach to the conceptual specification of information systems. Finally, Section 9 contains concluding remarks.

This work was partly supported by CNPq, under grants 473110/2008-3, 313031/2009-6 and 557128/2009-9, by FAPERJ under grant E-26/170028/2008, and by CAPES under grant CAPES/PROCAD NF 21/2009.

2. RELATIONAL MODEL

Contributions to the development of the relational model can be traced 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 while providing the advantages of a set-processing capability.

In a second early paper [Furtado 1978], the relational model was studied from three interdependent viewpoints. Relational databases were first modeled by directed hypergraphs, a concept derived in a straightforward way from Berge's hypergraph theory. Then, the abstract directed hypergraphs were interpreted using a linguistic model. Finally, the hypergraphs were represented with the help of relations and additional structures. Normalization was then discussed in the context of the three approaches.

A variant of Dynamic Logic was introduced in [Casanova and Bernstein 1980] to formalize a relational database description language (DDL) and a data manipulation language (DML). 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.

Recall that the original theory of data dependencies 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. 1984]. 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 14^{th} and 15^{th} VLDB Conferences [Casanova et al. 1988; Casanova et al. 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 1984] 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 1^{st} 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 2^{nd} 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 8^{th} 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 9^{th} ER Conference, two papers were presented. The first paper [Furtado and Casanova 1990b] 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 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 15^{th} 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, summarized work on mapping entity-relationship schemas into 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 28^{th} 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 1984; 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 7^{th} 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 3^{rd} PODS Conference [Casanova et al. 1984]. 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

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 10^{th} 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. 1991]. 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 [Furtado and Casanova 1990a].

These ideas were expanded later on into a model of question-answering [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. 1994] 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 determine 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 were integrated through the application of a plan-recognition / plan-generation paradigm [Furtado and Ciarlini 1997; 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 [Wilensky 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 were provided for planning and for the simulated execution of the generated plans [Ciarlini and Furtado 1999a; 1999b; 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].

The application of the plan-recognition / plan-generation paradigm to the narrative domain [Furtado 2004] was presented at the XIX Brazilian Symposium on Data Bases as an invited talk, on which occasion the author received a prize from the Brazilian Computer Society, acknowledging his contributions to database research.

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 and on the problem of publishing Linked Data. Some of the early 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] expressed our position with respect to the most frequent approaches to the schema matching problem, that is, the problem of defining concepts of a target schema T in terms of the concepts of a source schema S. The main points can be summarized as follows.

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.

The problem of matching two schemas that belong to an OWL dialect was addressed in [Leme et al. 2009; 2010], following an instance-based approach. The schema matching problem was 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 was 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 were extended in [Casanova et al. 2012] 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. 1984].

We now 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.

Finally, we briefly mention recent work on Linked Data. We first observe that disciplined schema design is one of the basic ideas behind Linked Data, and it should therefore be supported by triplification tools, that is, tools that publish databases as sets of RDF triples. Indeed, a major step of the triplification process is deciding how to represent database schema concepts in terms of an RDF vocabulary. The construction of this vocabulary is extremely important, because the more one reuses well-known standard vocabularies, the easier it will be to interlink the resulting triples to other existing datasets.

Linked Data is in fact the standard that W3C recommends for publishing Open Government Data (OGD). The United States of America and the United Kingdom released their government data portals in 2009 and 2010, respectively. In September 2011, Brazil became a member of the Open Government Partnership, a multinational initiative to promote worldwide adoption of OGD. The commitment includes a presidential mandate for the launch of the Brazilian Open Government Data portal, *DadosGov* (www.dados.gov.br).

In [Salas et al. 2010], we described the StdTrip tool, a triplification tool that facilitates the reuse of standard, W3C recommended RDF vocabularies or otherwise suggests the reuse of vocabularies already adopted by other RDF datasets. In [Salas et al. 2011], we discussed how to use the tool in the context of publishing Open Government Data. Finally, in [Breitman et al. 2011], we presented a concrete experience with government data mashups, including data from the DadosGov portal.

Future research will address methods, tools and experiments related to the description, publication and consumption of Linked Data sets. As for the description of Linked Data sets, the research will emphasize the quality of the final set of RDF triples, measured by the ability to identify their semantics, via strategies to generate integrity constraints as well as URIs. On what concerns the publication of Linked Data, the research will expand the traditional triplication process to include the creation of Web pages with embedded RDFa. Research on the consumption of Linked Data will explore SPARQL query optimization methods in the context of exploratory as well as mediated processing. It will also address strategies to create Linked Data mashups. Finally, the research will explore cloud computing platforms to store, index and consume large sets of Linked Data.

8. SEMIOTIC APPROACH

Based on Kenneth Burke's discussion [Burke 1969] of the so-called four master tropes - metonymy, metaphor, irony and synecdoche -, we identified four types of semiotic relations that can exist not only between facts, but also between events and between agents, which we denominated, respectively: syntagmatic, paradigmatic, antithetic and meronymic relations. Informally speaking, syntagmatic relations refer to connectivity, paradigmatic relations to similarity and analogy, antithetic relations to negation, and meronymic relations to hierarchy.

The paradigmatic relations were the first to attract our attention. In a SIGPLAN Notices paper [Furtado 1992], belonging to logic programming rather than to the database area, we argued that a powerful kind of reasoning by analogy is provided by combining unification with most specific generalization. We presented a revised version of an existing algorithm to compute the most specific generalization of terms, which decides correctly whether or not new variables should be introduced in each case. We also provided programs to perform unification and most specific generalization over frames, a data structure which would be of major importance for the practical application of our ideas, as will be repeatedly stressed in the sequel.

Our paper presented at SBBD 2007 [Barbosa et al. 2007] can be regarded as a first attempt to deal with paradigmatic relations in the context of databases. The motivating problem was that databases, particularly when storing heterogeneous, sparse semi-structured data, tend to provide incomplete information and information which is difficult to categorize. The paper first considers how to classify entity instances as members of entity classes organized in a lattice-like generalization/specialization hierarchy. Then, it describes how the frame representation employed for instances and classes, as well as the closeness criterion involved in the classification method, favors the practical use of similarity and analogy, where similarity refers to instances within the same class, and analogy involves different classes. Finally, the paper argues that similarity and analogy facilitate querying semi-structured data.

A more in-depth investigation of classification methods based on frames was the object of a more recent work [Nunes and Casanova 2010]. In fact, the problem of data classification goes back to the definition of taxonomies covering knowledge areas. With the advent of the Web, the amount of data available increased several orders of magnitude, making manual data classification impossible. The paper presents a tool to automatically classify semi-structured data, represented by frames, without any previous knowledge about structured classes. The tool uses a variation of the K-Medoid algorithm and organizes a set of frames into classes, structured as a strict hierarchy.

The next step, still focusing on paradigmatic relations and the corresponding trope, metaphor, was to promote a reuse strategy, whereby new conceptual specifications might be partly derived from previous ones. A paper along this line was presented at CIKM [Breitman et al. 2007]. Metaphor is not merely a rhetorical device, characteristic of language alone, but rather a fundamental feature of the human conceptual system. A metaphor is understood by finding an analogy mapping between two domains. The paper argued that analogy mappings facilitate conceptual modeling by allowing the designer to reinterpret fragments of familiar conceptual models in other contexts. The contributions of the paper were expressed within the tradition of the ER model, the Description Logic framework and as extensions of the OWL.

This reuse strategy was further examined in [Breitman et al. 2007; Furtado et al. 2008]. These papers argued in favor of a database conceptual schema and Semantic Web ontology design discipline that explores analogy mappings to reuse the structure and integrity constraints of conceptual models, stored in a repository. We presupposed that a team of expert conceptual designers would build a standard repository of source conceptual models, which less experienced designers would use to create new target conceptual models in other domains. The target models will then borrow the structure and the integrity constraints from the source models by analogy. The concepts were once again expressed in the contexts of Description Logics, the RDF model and OWL to reinforce the basic principles and explore additional questions, such as the consistency of the target model.

Reusing a conceptual schema is of course a multi-phase process. After finding a suitable source schema, adaptations will often be needed in view of conflicts with the target schema being designed. The notion of *blending* [Fauconnier and Turner 1994] was exploited for this objective in [Casanova et al. 2008]. To support the generation of database schemas of information systems, the paper proposed a five-step design process that explores the notions of generic and blended spaces and favors the reuse of predefined schemas. The use of generic and blended spaces is essential to achieve the passage from the source space into the target space in such a way that differences and conflicts can be detected and, whenever possible, conciliated. The convenience of working with multiple source schemas to cover distinct aspects of a target schema, as well the possibility of creating schemas at the generic and blended spaces, were also considered. Notice that, as we would indicate more explicitly in later articles, the presence of conflicts already suggests the need to deal with antithetic relations.

As mentioned before, since from our already referred SBBD paper [Barbosa et al. 2007], we have been using frames and frame-sets as a more flexible data structure than relational tuples and tables. At the 27^{th} ER Conference [Furtado et al. 2008], as we proceeded to show how to extend the reuse strategy to the design of dynamic schemas, we employed *plots*, also defined as a frame-like data structure. A plot is a partially ordered set of events. Plot analysis is a relevant source of knowledge about the agents' behavior when accessing data stored in the database. It relies on logical logs which register the actions of individual agents. The paper proposed techniques to analyze and reuse plots based on the concepts of similarity and analogy. The concept of similarity was applied to organize plots as a library, and to explore the reuse of plots in the same domain. By contrast, the concept of analogy helps reuse plots across different domains. The techniques proposed in the paper find applications in areas such as digital storytelling and emergency response information system, as well as some traditional business applications.

Our first study wherein all four semiotic relations were explicitly discussed was indeed presented at a digital storytelling conference, namely SBGames [Ciarlini et al. 2008]. In that paper, the process of plot composition in the context of interactive storytelling was considered under a fourfold perspective, in view of syntagmatic, paradigmatic, antithetic and meronymic relations between the constituent events. These relations were then shown to be associated with the four major tropes of semiotic research. A conceptual model and set of facilities for interactive plot composition and adaptation dealing with the four relations was described. To accommodate antithetic relations, corresponding to the irony trope, our plan-based approach leaves room for the unplanned. A simple storyboarding prototype tool has been implemented to conduct experiments. In another paper [Barbosa et al. 2010], already mentioned in Section 6, we utilized the semiotic relations to characterize classes of characters (agents, in the context of business information systems) according to their mutually interfering behavior in decision-making processes.

As remarked earlier, frames and plots became increasingly important to our research projects. The ER model is arguably today's most widely accepted basis for the conceptual specification of information systems. A further common practice is to use the relational model at an intermediate logical stage, in order to adequately prepare for physical implementation. Although the relational model still works well in contexts relying on standard databases, it imposes certain restrictions, not inherent in ER specifications, which make it less suitable in Web environments. Our 28^{th} ER Conference invited paper [Furtado et al. 2009], mentioned at the end of Section 3, recommends frames as an alternative to move from ER specifications to logical stage modeling, and treats frames as an abstract data type equipped with a Frame Manipulation Algebra. It is argued that frames, with a long tradition in AI applications, are able to accommodate the irregularities of semi-structured data, and that frame-sets generalize relational tables, allowing to drop the strict homogeneity requirement. The paper includes examples to help describe the use of the operators.

Likewise, a Plot Manipulation Algebra was proposed to handle plots in an ICEC conference [Karlsson et al. 2009]. The seven basic operators, equally named in both the Frame Manipulation Algebra and in the Plot Manipulation Algebra, and working respectively on frames and plots, were introduced in view of the four fundamental semiotic relations, as indicated below:

—syntagmatic relations - product, projection

—paradigmatic relations - union, selection

—antithetic relations - difference

—meronymic relations - combination, factoring

The operators in the first three lines above encompass the equivalent to the five basic operators of Codd's relational algebra (product, projection, union, selection, difference), plus two operators (combination, factoring) to handle the hierarchical structures induced by the meronymic relations, which would correspond to non-first-normal form (NF2) relations in the relational model (cf. our algebra of quotient relations [Furtado and Kerschberg 1977]). Thus, it seems fair to claim that our algebras are *semiotically complete*, which covers an ampler scope than that of Codd's relational algebra. Prototype logic-programming tools have been developed to experiment with the Frame Manipulation Algebra and the Plot Manipulation Algebra.

The pragmatic aspects of information systems constitute the main thrust of our present work, strongly influenced by the fundamental semiotic concepts exposed in this section. At the Second Workshop of the Brazilian Institute for Web Science Research [Barbosa et al. 2011], we argued for this orientation, which becomes increasingly relevant with the transition from the closed world of the old proprietary databases to the open world of the Web. Our view of information systems recognizes that, in order to serve as a basis for an effective communication process, their conceptual specification is comparable to the definition of a specialized language. Accordingly, it must pass through four levels: lexical, syntactic, semantic, and pragmatic [Morris 1938]. 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.

This pragmatic orientation, as we soon realized, is fully consistent with our conceptual design method which, as gradually exposed in the preceding sections, 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. And our plan-recognition / plan-generation paradigm puts together all these aspects and leads to executable specifications, which allow simulation experiments to effectively test the usability of the proposed conceptual design.

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. Sections 2 to 6 presented a retrospective of, roughly, the first 25 years. Section 7 summarized current work on the Deep Web and on Linked Data. Section 8 discussed recent work that departed from the more familiar database research tradition, but it brought 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, [Breitman et al. 2007], surveying in depth various topics pertaining to the Semantic Web technology, such as ontologies, and [Casanova et al. 2005], which contains an in-depth presentation of geographic databases. As textbooks for undergraduate and introductory graduate courses, we started long ago with [Furtado and S.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 was stressed in Section 8.

A number of visiting researchers had a major influence on the early development of our group, among which we may cite C.C. Gotlieb, E.F. Codd, C.J. Date, M.M. Zloof, M.R. Stonebraker, E.J. Neuhold and R. Fagin. Larry Kerschberg, today at George Mason University, was one of our most active colleagues for several years. We are grateful to José Paulo Schiffini from IBM Brasil, who helped us organize the *Fifth International Conference on Very Large Data Bases*, held in Rio de Janeiro, on October 3-5, 1979.

We are also pleased to acknowledge our long standing collaboration with the database groups at the Department of Computer Science, UFMG, the Department of Computer Science, UFAM, the Department of Computing, UFC, the Institute of Computing, UNICAMP, the Department of Applied Informatics, UNIRIO, the Institute of Computing, UFF, the Image Processing Division, INPE, and

the multimedia group at the Department of Informatics, PUC-Rio, to name just a few. A large fraction of the references of this paper reflects their invaluable participation. As a further incentive, the INCT for Web Science is now offering a continuing opportunity for collaboration, especially on the topics related to Sections 7 and 8.

REFERENCES

Alcazar, J. J. P., Hemerly, A. S., Casanova, M. A., and Furtado, A. L. Cooperative interfaces for spatio-temporal databases. In *Proceedings of the International Conference on Systems Research, Informatics and Cybernetics, Focus Symposium on Database and Expert Systems*, 1994.

Barbosa, S. and Silva, B. In , *Interação Humano-Computador.* Campus/Elsevier, 2010.

Barbosa, S. D. J., Breitman, K. K., Furtado, A. L., and Casanova, M. A. Similarity and analogy over application domains. In *Simposio Brasileiro Banco Dados.* Joao Pessoa, Brazil, 2007.

Barbosa, S. D. J., Breitman, K. K., Furtado, A. L., and Casanova, M. A. The Semiotic Web. In *Proceedings of the Second Workshop of the Brazilian Institute for Web Science Research.* Rio de Janeiro, Brazil, 2011.

Barbosa, S. D. J., Furtado, A. L., and Casanova, M. A. A decision-making process for digital storytelling. In *Proceedings of the 2010 Brazilian Symposium on Games and Digital Entertainment.* Florianópolis, Brazil, pp. 1–11, 2010.

Borgida, A., Casanova, M. A., and Laender, A. H. F. pp. 1645–1649. In L. Liu and M. T. Ozsu (Eds.), *Logical Database Design: from Conceptual to Logical Schema.* Springer, pp. 1645–1649, 2009.

Brauner, D. F., Gazola, A., Casanova, M. A., and Breitman, K. K. Adaptative matching of database web services export schemas. In *Proceedings of the International Conference on Enterprise Information Systems.* pp. 49–56, 2008.

Breitman, K., Casanova, M. A., and Truszkowski, W. *Semantic Web: Concepts, Technologies and Applications (NASA Monographs in Systems and Software Engineering).* Springer-Verlag, 2007.

Breitman, K. K., Barbosa, S. D. J., Casanova, M. A., and Furtado, A. L. Conceptual modeling by analogy and metaphor. In *Proceedings of the ACM Conference on Information and Knowledge Management.* pp. 865–868, 2007.

Breitman, K. K., Barbosa, S. D. J., Casanova, M. A., Furtado, A. L., and Hinchey, M. G. Using analogy to promote conceptual modeling reuse. In *Proceedings of the Workshop on Leveraging Applications of Formal Methods, Verification and Validation.* pp. 111–122, 2007.

Breitman, K. K., Salas, P. E., Saraiva, D., Gama, V., Casanova, M. A., Viterbo, J., Franzosi, E., Pires, R., and Chaves, M. Integrating open government datasets to create mashups. Submitted for publication, 2011.

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

Casanova, M. *The Concurrency Control Problem for Database Systems.* Lecture Notes in Computer Science. Springer-Verlag, 1981.

Casanova, M., Camara, G., and Jr, C. A. D. *Bancos de dados geograficos.* MundoGEO, 2005.

Casanova, M. A. Schema matching revisited. In *Proceedings of the Brazilian Symposium on Data Bases.* pp. 380, 2007.

Casanova, M. A., Barbosa, S. D. J., Breitman, K. K., and Furtado, A. L. Generalization and blending in the generation of entity-relationship schemas by analogy. In *Proceedings of the International Conference on Enterprise Information Systems.* pp. 43–48, 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* 2 (3): 386–414, 1980.

Casanova, M. A., Breitman, K. K., Brauner, D. F., and Marins, A. L. A. Database conceptual schema matching. *Computer (Long Beach)* 40 (10): 102–104, 2007.

Casanova, M. A., Breitman, K. K., Furtado, A. L., Vidal, V. M. P., and Macedo, J. A. F. On the problem of matching database schemas. In *Conquering Complexity*, M. Hinchey and L. Coyle (Eds.). Springer London, pp. 431–461, 2012.

Casanova, M. A., Fagin, R., and Papadimitriou, C. H. Inclusion dependencies and their interaction with functional dependencies. *Journal of Computer & Systems Sciences International* 28 (1): 29–59, 1984.

Casanova, M. A. and Furtado, A. L. pp. 211–236. In , *On the Description of Database Transition Constraints using Temporal Logic.* Plenum Press, pp. 211–236, 1984.

Casanova, M. A., Furtado, A. L., and Tucherman, L. A software tool for modular database design. *ACM Transactions on Database Systems* 16 (2): 209–234, 1991.

Casanova, M. A., Lauschner, T., Leme, L. A. P. P., Breitman, K. K., Furtado, A. L., and Vidal, V. M. P. Revising the constraints of lightweight mediated schemas. *Data Knowl. Eng.* 69 (12): 1274–1301, 2010.

Casanova, M. A., Tucherman, L., and Furtado, A. L. Enforcing inclusion dependencies and referencial integrity. In *Proceedings of the International Conference on Very Large Data Bases.* pp. 38–49, 1988.

Casanova, M. A., Tucherman, L., Furtado, A. L., and Braga, A. P. Optimization of relational schemas containing inclusion dependencies. In *Proceedings of the International Conference on Very Large Data Bases*. pp. 317–325, 1989.

Casanova, M. A., Tucherman, L., and Laender, A. H. F. Algorithms for designing and maintaining optimized relational representations of entity-relationship schemas. In *Proceedings of the International Conference on Entity-Relationship Approach. Lausanne*. pp. 361–374, 1990.

Casanova, M. A., Tucherman, L., and Laender, A. H. F. On the design and maintenance of optimized relational representations of entity-relationship schemas. *Data & Knowledge Engineering* 11 (1): 1–20, 1993.

Casanova, M. A., Veloso, P. A. S., and Furtado, A. L. Database specification formalisms: An eclectic perspective. In *Proceedings of the ACM SIGACT-SIGMOD Symposium on Principles of Database Systems*. ACM, pp. 110–118, 1984.

Castilho, J. M. V., Casanova, M. A., and Furtado, A. L. A temporal framework for database specifications. In *Proceedings of the International Conference on Very Large Data Bases*. pp. 280–291, 1982.

Ciarlini, A. E. M., Barbosa, S. D. J., Casanova, M. A., and Furtado, A. L. vent relations in plot-based plot composition. In *Proceedings of the Brazilian Symposium on Computer Games and Digital Entertainment*. pp. 31–40, 2008.

Ciarlini, A. E. M., Casanova, M. A., Furtado, A. L., and Veloso, P. A. S. Modeling interactive storytelling genres as application domains. *Journal of Intelligent Information Systems* 35 (3): 31–40, 2010.

Ciarlini, A. E. M., Feijó, B., and Furtado, A. L. An integrated tool for modelling, generating and exhibiting narratives. In *Proceedings of the AI, Simulation and Planning in High Autonomy Systems*. pp. 150–154, 2002.

Ciarlini, A. E. M. and Furtado, A. L. Interactive multistage simulation of goal-driven agents. *Journal of the Brazilian Computer Society* 6 (2), 1999a.

Ciarlini, A. E. M. and Furtado, A. L. Simulating the interaction of database agents. In *Proceedings of the Database and Expert Systems Applications Conference*. pp. 499–510, 1999b.

Ciarlini, A. E. M. and Furtado, A. L. Understanding and simulating narratives in the context of information systems. In *Proceedings of the International Conference on Conceptual Modeling*. pp. 291–306, 2002.

Ciarlini, A. E. M. and Furtado, A. L. Towards a plan-based learning environment. In *Proceedings of the PGL Database Research Conference*, 2003.

Ciarlini, A. E. M., Pozzer, C. T., Furtado, A. L., and Feijó, B. A logic-based tool for interactive generation and dramatization of stories. In *Proceedings of the ACM-SIGCHI International Conference on Advances in Computer Entertainment Technology*. pp. 133–140, 2005.

Ciarlini, A. E. M., Veloso, P. A. S., and Furtado, A. L. A formal framework for modelling at the behavioural level. In *Information modeling and knowledge bases XII*. pp. 107–122, 2000.

Fauconnier, G. and Turner, M. Conceptual projection and middle spaces, 1994.

Fikes, R. and Nilsson, N. J. Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence* 2 (3/4): 189–208, 1971.

Furtado, A. and Casanova, M. An information system environment based on plan generation. In *Proceedings of the Working Conference on Cooperating Knowledge based Systems*, 1990a.

Furtado, A. L. Formal aspects of the relational model. *Information Systems* 3 (2): 131–140, 1978.

Furtado, A. L. Horizontal decomposition to improve a non-bcnf scheme. *ACM SIGMOD Record* 12 (1): 26–32, 1981.

Furtado, A. L. An informal approach to formal specifications. *ACM SIGMOD Record* 14 (1): 45–54, 1984.

Furtado, A. L. Analogy by generalization and the quest of the grail. *ACM SIGPLAN Notices* 27 (1): 105–113, 1992.

Furtado, A. L. Narratives and temporal databases: An interdisciplinary perspective. In *Conceptual modeling: historical perspectives and future directions*, 1999.

Furtado, A. L. Narratives over real-life and fictional domains. In *Proceedings of the Brazilian Symposium on Data Bases*. pp. 4–12, 2004.

Furtado, A. L. and Casanova, M. A. Updating relational views. In *Query Processing in Database Systems*. Springer, pp. 127–142, 1985.

Furtado, A. L. and Casanova, M. A. Plan and schedule generation over temporal databases. In *Proceedings of the International Conference on Entity-Relationship Approach*. pp. 235–248, 1990b.

Furtado, A. L., Casanova, M. A., Barbosa, S. D. J., and Breitman, K. K. Analysis and reuse of plots using similarity and analogy. In *Proceedings of the International Conference on Conceptual Modeling*. pp. 355–368, 2008.

Furtado, A. L., Casanova, M. A., Breitman, K. K., and Barbosa, S. D. J. A frame manipulation algebra for er logical stage modelling. In *Proceedings of the International Conference on Conceptual Modeling*. pp. 9–24, 2009.

Furtado, A. L. and Ciarlini, A. E. M. Plots of narratives over temporal databases. In *In Proceedings of the International Conference and Workshop on Database and Expert Systems Applications*. pp. 590–595, 1997.

Furtado, A. L. and Ciarlini, A. E. M. Operational characterization of genre in literary and real-life domains. In *Proceedings of the Conceptual Modeling Conference*. pp. 460–474, 1999.

Furtado, A. L. and Ciarlini, A. E. M. Generating narratives from plots using schema information. In *Proceedings of the International Workshop on Applications of Natural Language for Information Systems*. pp. 17–29, 2000a.

Furtado, A. L. and Ciarlini, A. E. M. In , *The plan recognition / plan generation paradigm*. Springer-Verlag, 2000b.

Furtado, A. L. and Ciarlini, A. E. M. Constructing libraries of typical plans. In *Proceedings of the Conference on Advanced Information Systems Engineering*. pp. 124–139, 2001.

Furtado, A. L. and Ciarlini, A. E. M. Cognitive and affective motivation in conceptual modelling. *Revista Colombiana de Computación* 3 (2), 2002.

Furtado, A. L. and Kerschberg, L. An algebra of quotient relations. In *Proceedings of the ACM SIGMOD international conference on Management of data*. pp. 1–8, 1977.

Furtado, A. L. and Maibaum, T. S. E. An informal approach to formal algebraic specifications. *Computer Journal* 28 (1): 59–67, 1985.

Furtado, A. L. and Neuhold, E. J. *Formal Techniques for Data Base Design*. Springer-Verlag New York, Inc., 1986.

Furtado, A. L., Sevcik, K. C., and dos Santos, C. S. Permitting updates through views of data bases. *Information Systems* 4 (4): 269–283, 1979.

Furtado, A. L. and S.Santos, C. *Organizacao de bancos de dados*. Campus, 1979.

Furtado, A. L., Veloso, P. A. S., and Castilho, J. M. V. Verification and testing of s-er representations. In *Proceedings of the International Conference on the Entity-Relationship Approach*. pp. 123–147, 1981.

Gomes, R. V., Leme, L. A. P. P., and A.Casanova, M. Matchmaking âĂŞ a tool to match owl schemas. *Revista de Informática Teórica e Aplicada* vol. 16, pp. 71–76, 2010.

Hemerly, A. S., Casanova, M. A., and Furtado, A. L. Cooperative behavior through request modification. In *Proceedings of the International Conference on Entity-Relationship Approach*. pp. 607–621, 1991.

Hemerly, A. S., Casanova, M. A., and Furtado, A. L. Avoiding misconstruals in database systems: A default logic approach. *IEEE Transactions on Knowledge and Data Engineering* 5 (6): 994–996, 1993.

Hemerly, A. S., Casanova, M. A., and Furtado, A. L. Nonstandard queries and nonstandard answers. Oxford University Press, Exploiting user models to avoid misconstruals, pp. 73–97, 1994.

Karlsson, B., Barbosa, S. D. J., Furtado, A. L., and Casanova, M. A. A plot-manipulation algebra to support digital storytelling. In *Proceedings of the International Conference on Entertainment Computing*. pp. 132–144, 2009.

Kautz, H. A. Reasoning about plans. Morgan Kaufmann Publishers Inc., A formal theory of plan recognition and its implementation, pp. 69–124, 1991.

Laender, A. H. F., Casanova, M. A., de Carvalho, A. P., and Ridolfi, L. F. G. G. M. An analysis of sql integrity constraints from an entity-relationship model perspective. *Information Systems* 19 (4): 331–358, 1994.

Leme, L. A. P. P., Brauner, D. F., Breitman, K. K., Casanova, M. A., and Gazola, A. Matching object catalogues. *Innovations in Systems and Software Engineering* 4 (4): 315–328, 2008.

Leme, L. A. P. P., Casanova, M. A., Breitman, K. K., and Furtado, A. L. Instance-based owl schema matching. In *Proceedings of the International Conference on Enterprise Information Systems*. pp. 14–26, 2009.

Leme, L. A. P. P., Casanova, M. A., Breitman, K. K., and Furtado, A. L. Owl schema matching. *Journal of the Brazilian Computer Society* 16 (1): 21–34, 2010.

Morris, C. *Foundations of the theory of signs*. University of Chicago Press, 1938.

Nunes, B. and Casanova, M. A. A frame-based system for automatic classification of semi- structured data. *Revista de InformÃątica Teórica e Aplicada* 16 (2): 87–92, 2010.

Piccinini, H., Lemos, M., Casanova, M. A., and Furtado, A. L. W-ray: A strategy to publish deep web geographic data. In *Proceedings of the International Workshop on Semantic and Conceptual Issues in GIS*. pp. 2–11, 2010.

Propp, V. *Morphology of the Folktale*. University of Texas Press, 1968.

Quine, W. V. Ontological relativity. *The Journal of Philosophy* 65 (7): 185–212, 1968.

Salas, P. E., Breitman, K. K., F., J. V., and Casanova, M. A. Interoperability by design using the stdtrip tool: an a priori approach. In *Proceedings of the International Conference on Semantic Systems*. pp. 97–105, 2010.

Salas, P. E., Viterbo, J., Breitman, K., and Casanova, M. A. Stdtrip: Promoting the reuse of standard vocabularies in open government data. In *Linking Government Data*, D. Wood (Ed.). Springer New York, pp. 113–134, 2011.

Santos, C. S., Maibaum, T. S. E., and Furtado, A. L. Conceptual modeling of data base operations. *nternational Journal of Computer and Information Science* vol. 10, pp. 299–314, 1981.

Santos, C. S., Neuhold, E. J., and Furtado, A. L. A data type approach to the entity-relationship model. In *Proceedings of the International Conference on the Entity-Relationship Approach to Systems Analysis and Design*. pp. 103–119, 1979.

Schiel, U., Furtado, A. L., Neuhold, E. J., and Casanova, M. A. Towards multi-level and modular conceptual schema specifications. *Information Systems* 9 (1): 43–57, 1984.

Sena, G. J. and Furtado, A. L. Towards a cooperative question-answering model. In *Proceedings of the International Conference on Flexible Query Answering Systems*. pp. 354–365, 1998.

Silva, A. S., Laender, A. H. F., and Casanova, M. A. An approach to maintaining optimized relational representations of entity-relationship schemas. In *Proceedings of the International Conference on Conceptual Modeling*. pp. 292–308, 1996.

Silva, A. S., Laender, A. H. F., and Casanova, M. A. On the relational representation of complex specialization structures. *nformation Systems* 25 (5): 399–415, 2000.

Tucherman, L., Casanova, M. A., and Furtado, A. L. The chris consultant-a tool for database design and rapid prototyping. *Information Systems* 15 (2): 187–195, 1990.

Tucherman, L., Casanova, M. A., Gualandi, P. M., and Braga, A. P. A proposal for formalizing and extending the generalization and subset abstractions in the enity-relationship model. In *Proceedings of the International Conference on Entity-Relationship Approach*. pp. 27–41, 1989.

Veloso, P. A. S., de Castilho, J. M. V., and Furtado, A. L. Systematic derivation of complementary specifications. In *Proceedings of the International Conference on Very Large Data Bases*. pp. 409–421, 1981.

Veloso, P. A. S. and Furtado, A. L. Towards simpler and yet complete formal specifications. In *Proceedings of the IFIP WG8.1 Working Conference*. pp. 174–188, 1985.

Vidal, V. M. P. and Casanova, M. A. Towards a sound view integration methodology. In *Proceedings of the ACM SIGACT-SIGMOD Symposium on Principles of Database Systems*. pp. 36–47, 1983.

Wilensky, R. *Planning and understanding: a computational approach to human reasoning*. Addison-Wesley Pub. Co. Advanced Book Program, 1983.