research@cse

A compilation of research abstracts

Department of Computer Science and Engineering
Indian Institute of Technology Bombay
March, 2015
Message from the Head

Welcome to the Research Scholar Poster Mela. The department of Computer Science and Engineering at IIT Bombay has had an impressive history of leading computer science efforts in India, a very good present, positively impacting computer science research internationally, and we believe it will have a great future as our research programs grow and improve. And the major factor in our future are our research scholars, who have grown in both numbers and in quality over the past years.

Today our department has internationally recognized groups in many areas. To name a few in alphabetical order: algorithms and complexity, data mining/machine learning, database systems, graphics, formal methods, information retrieval, natural language processing, and networks/distributed systems, as well as other smaller groups.

Our research work is today published in top tier (A/A+ level) conferences and journals, and many papers from IIT Bombay have been highly cited. Our faculty have received international recognition, as evidenced by the number of faculty who are/have been program chairs of top-tier conferences, and/or editors of top-tier journals.

Our research scholars have also won international recognition, winning fellowships and awards from leading companies such as Microsoft, Yahoo, IBM, TCS and Infosys, to name a few, with several of the awards coming from international competitions. The first ACM India doctoral dissertation award was won by Ruta Mehta in 2012.

But we must continue to aim higher. Our goal is to be recognized as not just the best or one of the best in India, but as one of the best overall in the world. With the efforts of our faculty and research scholars, supported by our masters students as well as bachelors students, I believe we can join the top 10 to 20 rank internationally in CSE within, say, a decades time.

So let me sign off, with kudos to our research scholars for their achievements, which we seek to showcase at the mela, and with best wishes for continued research success.

S. Sudarshan
## Contents

A Memory-access Efficient Parallel Algorithm for Spectrum-based Short Read Error Correction  
Sriram Ponnambalam C Srinivas Aluru .......................... 1

Harnessing eye-tracking to understand sentiment annotation  
Aditya Joshi Abhijit Mishra Pushpak Bhattacharyya 2

Addressing Class Imbalance in Grammatical Error Detection  
Anoop Kundukuttan Pushpak Bhattacharyya ................. 4

Coreference Resolution for Better Information Retrieval from Indian Classical Music Forums  
Joe Cheri Ross Prof. Pushpak Bhattacharyya .............. 6

Multilingual Multiword Expressions Extraction for Indian Languages  
Rudra Murthy V Pushpak Bhattacharyya ....................... 7

Lexico-Syntactic Simplification of Text  
Sandeep Mathias Pushpak Bhattacharyya ....................... 8

Unsupervised Most Frequent Sense Detection using Word Embeddings  
Sudha Bhingardive Pushpak Bhattacharyya ................. 9
Bounded-Rate Multi-Mode Systems Based
Motion Planning
Devendra Bhave Sagar Jha Shankara Narayanan
Krishna Ashutosh Trivedi
Collaborator(s): Sven Schewe

Bounded-rate multi-mode systems are hybrid systems that can switch among a finite set of modes. Its dynamics is specified by a finite number of real-valued variables with mode-dependent rates that can vary within given bounded sets. Given an arbitrary piecewise linear trajectory, we study the problem of following the trajectory with arbitrary precision, using motion primitives given as bounded-rate multi-mode systems. We give an algorithm to solve the problem and show that the problem is co-NP complete. We further prove that the problem can be solved in polynomial time for multi-mode systems with fixed dimension. We study the problem with dwell-time requirement and show the decidability of the problem under certain positivity restriction on the rate vectors. Finally, we show that introducing structure to the multi-mode systems leads to undecidability, even when using only a single clock variable.

Bibliography

may include imperative code as well as SQL queries. Our techniques can be used for performing optimizations that holistic optimization techniques proposed earlier cannot perform, including rewriting of (some) imperative code into SQL. Equally importantly, the algebraic representation of an application can be used as an input for other tasks such as keyword search on forms, and test data generation.

Bibliography


Traditionally, database applications have been optimized independently on two fronts - source code optimizations by the language compiler and query optimization at the database. These techniques, however, do not guarantee the global optimum, as they fail to recognize opportunities for optimization that lie hidden, both from the database query optimizer and the programming language compiler, working in isolation. A holistic view of the database application as a single unit is necessary to perform such optimizations.

Over the last decade, there have been many interesting developments in optimizing database applications. Techniques from database query optimization, program analysis, and program synthesis have been adopted to solve various problems in the area of holistic optimization. Guravannavar et al. [1] propose to alleviate this problem by batching, i.e., replacing multiple calls to a query by a batched call to a correspondingly rewritten query, which allows the use of set-oriented query execution plans. Ramachandra et al. [2] address this problem through asynchronous query submission. Since asynchronous calls are non-blocking, they allow applications to reduce the perceived latency by overlapping program execution with query retrieval. Ramachandra et al. [3] propose a technique called query anticipability analysis, based on the data flow framework anticipable expressions analysis, to determine the earliest program point at which a prefetch instruction for a query can be inserted, without any prefetches being wasted. Simhadri et al. [4] proposed techniques to algebraize imperative constructs in user defined functions (UDFs), to extract a single relational algebra expression for the entire UDF body, so that for a given query, any performance bottlenecks can be identified and rectified.

Currently, we are working on techniques to get a concise algebraic representation of (parts of) a database application, which

\begin{itemize}
  \item \textbf{K Venkatesh Emani S Sudarshan}
\end{itemize}

\section*{Holistic Optimization of Database Applications}

\begin{itemize}
  \item Collaborator(s): Nagakishore Jammula
\end{itemize}

\section*{Bibliography}


DNA Read Error Correction

\begin{itemize}
  \item \textbf{Sriram Ponnambalam C Srinivas Aluru}
\end{itemize}

\section*{Algorithm for Spectrum-based Short Read Error Correction}

A Memory-access Efficient Parallel Algorithm for Spectrum-based Short Read Error Correction

\begin{itemize}
  \item Collaborator(s): Nagakishore Jammula
\end{itemize}

\section*{DNA Read Error Correction}

\begin{itemize}
  \item \textbf{Sriram Ponnambalam C Srinivas Aluru}
\end{itemize}
The XDa-TA System for Automated Grading of SQL Query Assignments

Bikash Chandra S. Sudarshan

Collaborator(s): Amol Bhangdiya, Biplab Kar, Bharath Radhakrishnan, K. V. Maheshwara Reddy, Shetal Shah

Grading of student SQL queries is usually done by executing the query on sample datasets (which may be unable to catch many errors) and/or by manually comparing/checking a student query with the correct query (which can be tedious and error prone). In this demonstration we present the XDa-TA system which can be used by instructors and TAs for grading SQL query assignments automatically. Given one or more correct queries for an SQL assignment, the tool uses the XData system to automatically generate datasets that are designed specifically to catch common errors. The grading is then done by comparing the results of student queries with those of the correct queries against these generated datasets; instructors can optionally provide additional datasets for testing. The tool can also be used in a learning mode by students, where it can provide immediate feedback with hints explaining possible reasons for erroneous output. This tool could be of great value to instructors particularly, to instructors of MOOCs.

Harnessing eye-tracking to understand sentiment annotation

Aditya Joshi Abhijit Mishra Pushpak Bhattacharyya

Sentiment analysis is the task of predicting opinion in text. Typical supervised approaches rely on datasets annotated with sentiment. The process of assigning sentiment to a piece of text is called sentiment annotation. In this poster, we describe two of our works that aim to get insights from cognitive technologies relying on eye-tracking, in order to understand sentiment annotation. We use our understanding for two separate tasks.

The first task is to measure difficulty of sentiment annotation. The effort required for a human to detect sentiment is not uniform for all texts. Our metric called “Sentiment Annotation Complexity (SAC)” [1] quantifies this effort, using linguistic properties of text. We use an eye-tracking device to record eye-movements during sentiment annotation. We then present a classifier to predict this metric for new text. This metric is useful for fine-tuned pricing - where a new labeled dataset can be built with an optimal budget based on how difficult each of the document is, for sentiment annotation.

The second task is a cognitive study of sentiment annotation in humans [2]. Sentiment classification is often preceded by subjectivity classification. We wish to understand “do humans perform subjectivity extraction when reading for sentiment”. We observe a correlation between human behaviour and the way sentiment changes across a document. For documents containing ‘linear sentiment’, human annotators perform subjectivity extraction through anticipation, while for documents containing ‘oscillating sentiment’, they exhibit subjectivity extraction through homing.

Bibliography

Bibliography


Precise Call Disambiguation in the Presence of Function Pointers
Swati Rathi  Uday Khedker

Interprocedural data flow analysis increases the precision of data flow information by incorporating the effect of callers on callees and vice-versa. Thus it requires as input, the caller-callee relationships. In presence of calls through function pointers, discovering exact caller-callee relationship may not be possible at compile time and hence they have to be approximated. Safety of interprocedural analysis requires that caller-callee relationships should not be under-approximated, its precision requires that these relationships should not be over-approximated. Also if we over-approximate the callee information, we will have to consider additional callees. This will not only compute imprecise results but also affects the efficiency of the analysis. Hence, precise call disambiguation is essential for precise and efficient interprocedural analysis.

The caller-callee relationship is represented in the form of a call-graph. Many approaches for call graph construction exists in the literature.[1, 2, 3, 4] However, none of these approaches precisely disambiguates calls in the presence of function pointers. In the absence of function pointers, we traverse over the supergraph using call-strings based approach. [5] Call-strings helps us identify the valid caller to return back to in the appropriate context. Thus eliminating interprocedurally invalid paths. In the presence of function pointers, to eliminate interprocedurally invalid paths, we need to identify the valid callees to be invoked in the appropriate context. Thus the existing traversal mechanism is not sufficient. We thus propose an approach to redefine context to facilitate a precise traversal mechanism over the supergraph.

Addressing Class Imbalance in Grammatical Error Detection
Anoop Kunchukuttan  Pushpak Bhattacharyya

We address the challenge of class imbalance in the Grammatical error detection (GED) task for non-native speaker text. In such texts, the ratio of erroneous (positive) examples to error-free (negative) examples is very low. If the error detection problem is posed as a classification problem, this results in a high imbalance ratio among the classes. Common learning algorithms can achieve high training accuracy simply by labelling all examples as correct. Such a classifier is not useful for the task at hand.

Of course, sampling techniques (under, over and weighted) address this problem of learning in the presence of data skew [1]. We compare traditional sampling-based methods for addressing class imbalance on three GED tasks: noun number, article and preposition error detection. We show that these methods have limitations when applied to GED tasks because of the sparse, large feature space, uncertain sampling rate and the high imbalance ratio.

We show that learning classifiers by optimizing recall-oriented measures like the $F_2$ score delivers well-rounded, satisfactory performance in the following sense. For users like copy-editors, who detect and correct grammatical errors, the GED system should ideally catch all the errors, even if the precision is sacrificed to a certain extent. GED is thus a high recall demanding task. We attain this goal by directly optimizing $F_2$ while maintaining accuracy and precision above what is achieved through sampling methods. We also show that there are inherent difficulties in optimizing precision-oriented evaluation metrics like $F_{0.5}$.

The $F_\beta$ evaluation metrics do not decompose over the training examples. Hence, we use a max-margin formulation [2] based on StructSVM [3] to optimize such performance measures for binary classification problems. Exact and efficient optimization can be done for metrics that can be computed from the contingency table of a classifier’s results. e.g. $F_\beta$, recall, precision.
Study shows that the most common challenge faced by novice C/C++ programmers is the management of dynamic memory (heap). Understanding the concept of pointers in itself is nontrivial. Without this understanding, poorly written programs have memory leaks that impact the performance of the programs. Such programs use unnecessarily large system resources, and worst of all, they fail due to out-of-resource problems. As a consequence, analysis of heap memory is becoming increasingly important.

Heap data is potentially unbounded and seemingly arbitrary. We provide a high-level view of the abstraction techniques used for static analyses of heap data. We view a heap abstraction as consisting of two features: a heap model to represent the heap memory and a summarization technique for bounding the heap representation. We classify the models as storeless, store-based, and hybrid. We review various summarization techniques, viz. k-limiting summarization, allocation-site based summarization, pattern based summarization, variable based summarization, and other generic predicates based summarization.

We have studied how these summarization techniques have been adopted in the literature for different heap models. As program analysts, we are still facing the challenge of creating summaries that yield results that are both scalable to large sized programs and precise enough to produce useful results.
Flow and Context Sensitive Points-to Analysis using Higher Order Reachability

Pritam Gharat  Uday Khedker

The bottom up interprocedural approaches which construct summary flow functions for procedures and use them in the place of calls, have been effectively used for many analyses except for flow and context sensitive points-to analysis. The main difficulty in this is representing indirect accesses of pointees defined in the calling contexts. This is handled by constructing flow functions using placeholders and creating customized call-specific versions. However, their sizes are not bounded by the number of pointer variables.

We propose a bounded representation of summary flow functions called the higher order reachability graph (HRG) for points-to analysis that avoids explicit placeholders. HRG generalizes the use of graph reachability in program analysis to represent a dependence between variables by defining the notion of higher order reachability which relates (transitively indirect) pointees of a variable with those of another variable. HRGs for points-to analysis can distinguish between may and must points-to information thereby allowing strong updates.

Our empirical measurements on SPEC benchmarks show that the summary flow functions for most procedures are small and are used multiple times. For programs upto 40 kLoC, flow and context sensitive points-to analysis using HRGs is very efficient. Thus this is a promising direction for further investigations in efficiency and scalability of points-to analysis without compromising on precision.

Coreference Resolution for Better Information Retrieval from Indian Classical Music Forums

Joe Cheri Ross  Prof. Pushpak Bhattacharyya

Efficient music information retrieval (MIR) require to have meta information about music along with content based information in the knowledge base. Discussion forums on music are rich sources of information gathered from a wider audience. Taking into consideration the nature of text in these web resources, the yield of relation extraction is quite dependent on resolving the references in the document. Among the few music forums dealing with Indian classical music, rasikas.org[1] having rich information about artistes, raga and other music concepts is taken for our study. The forum posts generally contain anaphoric references to the main topic of the thread or any other entity in the discourse. We focus on features for coreference resolution for short discourse noisy text like that of forum posts. Since grammatical roles capture relation between mentions in a discourse, those features extracted from dependency parsing are widely explored along with semantic compatibility feature. On investigation of issues, the need for integrating known rules to the model emerged. Since Bayesian belief network is capable of integrating prior knowledge, mention pair classification is evaluated with Bayesian network. Modifications are to be proposed for coreference clustering to alleviate false alarms due to poor clustering.

Bibliography

[1] @ONLINE http://www.rasikas.org
perform the actual transportation. We claim that this schema-level treatment before migration makes the migration mechanism less expensive computationally, as compared to the existing methods of solving the same problem such as using the Petri net state equation \[1\]. Moreover, with the recent trends of pattern-based block-structured process models \[3\], verifying the schema compatibility is more intuitive and computationally cheap as compared to the methods presented in \[4\].

**Bibliography**


Catalog-based Token Transportation for Dynamic Workflow Evolution
Ahana Pradhan Rushikesh K. Joshi

Organizational goals are realized by executing business processes that involve people, resources, schedules and technology. Traditional workflow management systems (WFMS) are well-suited for rigid processes. However, the volatile nature of business processes requires intricate facilities for changing the workflows at runtime in WFMS in a valid and consistent manner. In absence of this support, an information system susceptible to changes needs to be tackled by slower porting processes, it not being immediately usable due to the not easily bridgeable gap between the pre-planned and the evolved actuality.

An evolutionary change includes process change at schema level and also instance migration for all running cases to the changed schema. The poster presents an algorithm called the Yo-Yo algorithm for Petri net models of workflows to carry out consistent runtime instance migration in this context. Figure 1a shows the schema of an imaginary workflow modeled in WF-net[2] and the runtime state of its instance [1], which is subjected to dynamic migration. The state in the new workflow (cf. Figure 1b) shows the computed state from which the execution should resume after the migration. The presented Yo-Yo algorithm computes this runtime state of the new process given the schemas of the two workflows and the old process state. The consistency on which the state-mappings are based on is the equality between the sets of completed activities in both of the processes.

The uniqueness of the Yo-Yo algorithm is that it is based on a catalog, which is an inventory of pre-computed state-mappings on the basis of consistency. The algorithm makes use of the cataloged mappings to carry out modular token transports as the migration pairs are folded and unfolded. This approach of token transportation first provides a schema-level treatment to the migration pair by establishing Yo-Yo compatibility, i.e. compatibility of folding hand-in-hand between the two schemas before and the evolved actuality.

Lexico-Syntactic Simplification of Text
Sandeep Mathias Pushpak Bhattacharyya

Simplification of text is the process in which text is simplified in such a way that more people can understand it without much difficulty. We all know that different types of text pose different problems to different readers. Young readers may not be able to read text with difficult words, while second language learners (and at times, native speakers as well) may struggle with text containing long sentences.

There are 2 types of simplification - Syntactic and Lexical simplification. Syntactic simplification involves splitting a sentence into one or more smaller sentences. This can be done by parsing the larger sentence, matching it with a rule for splitting, and splitting the sentence appropriately[1]. Lexical simplification involves substituting words or phrases with simpler words and phrases[2].

Both types of simplification have their own challenges. Syntactic Simplification has difficulties in maintaining cohesion in the document, ordering of sentences, preserving coreference links in the document, etc. Lexical simplification suffers from context sensitivity, as well as choosing if, and which, word (or phrase) has to be substituted, without changing the meaning of the document.

Bibliography
Architecture-Centric Software Evolution towards Generalisation

Suren Duhan Rushikesh K. Joshi

Software evolution requires changes in its architecture for many software. There are many requirements when a similar architecture is used to produce multiple software with a very few differences in their behaviour. When there are requirements of multiple instances of a software with some difference in the behaviour of the instances, then different versions of the same software are produced. This, most of the times, requires change in the code and/or architecture and causes the repetition of code or consumes more time. This modification takes time which further delays in addition of new features to the system (if needed and planned). We are proposing a way of generalising some such architectures where multiple software have some common behaviour. The generic architecture produced from them will be capable of reproducing all of the specific architectures with minimal changes required. Our approach uses Labelled Transition System (LTS) to extract common configurations of the specific architectures with minimal differences. Our aim is to automate the process of generalisation and can be easily ported to various domains.

Unsupervised Most Frequent Sense Detection using Word Embeddings

Sudha Bhingardive Pushpak Bhattacharyya

An acid test for any new Word Sense Disambiguation (WSD) algorithm is its performance against the Most Frequent Sense (MFS) baseline. The field of WSD has found the MFS baseline very hard to beat. Clearly, if researchers had access to MFS values, their struggle to better this heuristic will push the WSD frontier. However, getting MFS values requires the sense tagged Corpus (SemCor) which contains millions of sentences with each word annotated with all its senses. In this paper, we propose an unsupervised method for the MFS detection from the untagged corpora, which exploits word embeddings. We compare the word embeddings of a word with all its sense embeddings and obtain the predominant sense with the highest similarity. We observe the significant performance gain for Hindi WSD over the WordNet First Sense (WFS) baseline. However, the performance of our approach is language independent and can be easily ported to various domains.

An acid test for any new Word Sense Disambiguation (WSD) algorithm is its performance against the Most Frequent Sense (MFS) baseline. The field of WSD has found the MFS baseline very hard to beat. Clearly, if researchers had access to MFS values, their struggle to better this heuristic will push the WSD frontier. However, getting MFS values requires the sense tagged Corpus (SemCor) which contains millions of sentences with each word annotated with all its senses. In this paper, we propose an unsupervised method for the MFS detection from the untagged corpora, which exploits word embeddings. We compare the word embeddings of a word with all its sense embeddings and obtain the predominant sense with the highest similarity. We observe the significant performance gain for Hindi WSD over the WordNet First Sense (WFS) baseline. However, the performance of our approach is language independent and can be easily ported to various domains.

Unsupervised Most Frequent Sense Detection using Word Embeddings

Sudha Bhingardive Pushpak Bhattacharyya

An acid test for any new Word Sense Disambiguation (WSD) algorithm is its performance against the Most Frequent Sense (MFS) baseline. The field of WSD has found the MFS baseline very hard to beat. Clearly, if researchers had access to MFS values, their struggle to better this heuristic will push the WSD frontier. However, getting MFS values requires the sense tagged Corpus (SemCor) which contains millions of sentences with each word annotated with all its senses. In this paper, we propose an unsupervised method for the MFS detection from the untagged corpora, which exploits word embeddings. We compare the word embeddings of a word with all its sense embeddings and obtain the predominant sense with the highest similarity. We observe the significant performance gain for Hindi WSD over the WordNet First Sense (WFS) baseline. However, the performance of our approach is language independent and can be easily ported to various domains.
Formal methods for analysis of biological systems: A Boolean satisfiability approach

Sukanya Basu  Harshit Pande  Supratik Chakraborty  Akshay S

Biologists perform various wet-lab experiments and make certain observations. If the observations are interesting to them, they want to know in detail the mechanism by which the particular phenomenon is taking place. Obtaining this information solely by extensive experimentation is very expensive and often infeasible.

There are many databases, drawn by experts, that host several networks (pathways) explaining observed or well-known biological processes. They document in detail the interaction between several entities involved in that process. However, often there is no one model that can capture an observation made through lab experiments. It is also manually tedious to sieve through all such models to arrive at a hypotheses.

We strive to harness the power of these databases, by considering several network models and trying to explain a particular experimental profile. We try to use Boolean satisfiability checking, a well known and highly scalable formal method to perform this task. The poster attempts to put forth our ideas and approach, while also mentioning the various challenges faced, in trying to bring together two very divergent paradigms of biological sciences and computational methods.

A Generalization of the Łoś-Tarski Preservation Theorem

Abhisekh Sankaran  Bharat Adsul  Supratik Chakraborty

Preservation theorems in First Order (FO) logic have been extensively studied in model theory. A classical preservation theorem (also one of the earliest) is the Łoś-Tarski theorem, which gives semantic characterizations, over the class of all structures, of the \( \forall^* \) and \( \exists^* \) prefix classes\(^2\) of FO sentences, in terms of the properties of preservation under substructures and preservation under extensions, respectively. We have identified [1] new parameterized preservation properties that provide, for each natural number \( k \) as a parameter, semantic characterizations of the \( \exists^k \forall^* \) and \( \forall^k \exists^* \) prefix classes of FO sentences, over the class of all structures. These properties, which we call as preservation under substructures modulo \( k \)-cruxes and preservation under \( k \)-ary covered extensions respectively, correspond exactly to preservation under substructures and preservation under extensions, when \( k \) equals 0. As a consequence, we get a parameterized generalization of the Łoś-Tarski preservation theorem for sentences, in both forms of the latter, substructural and extensional. We call our characterizations collectively as the generalized Łoś-Tarski theorem for sentences.

To the best of our knowledge, our characterizations are the first to relate counts of quantifiers in (the leading block of) \( \exists^k \forall^* \) and \( \forall^k \exists^* \) sentences, to natural model-theoretic properties. The literature contains various characterizations of the \( \exists^k \forall^* \) and \( \forall^k \exists^* \) prefix classes in terms of preservation properties, examples of

\(^2\)The class \( \forall^* \) is the class of all FO sentences of the form \( \forall y_1 \ldots \forall y_n \psi(y_1, \ldots, y_n) \) where \( n \) is some natural number, and \( \psi \) is a quantifier-free formula with free variables \( y_1, \ldots, y_n \). The class \( \exists^* \) is defined similarly. Likewise, for a given natural number \( k \), the class \( \exists^k \forall^* \) is the class of all FO sentences of the form \( \exists x_1 \ldots \exists x_k \forall y_1 \ldots \forall y_n \psi(x_1, \ldots, x_k, y_1, \ldots, y_n) \) where \( n \) is some natural number, and \( \psi \) is a quantifier-free formula with free variables \( x_1, \ldots, x_k, y_1, \ldots, y_n \). The classes \( \forall^k \exists^* \), \( \exists^k \forall^* \) and \( \forall^* \exists^k \) are defined similarly.

the latter being those defined using ascending chains, descending chains or 1-sandwiches. None of these however relates the counts of quantifiers to any model-theoretic properties. Our characterizations are therefore finer than these characterizations of the literature.

Furthermore, unlike any of the aforementioned preservation properties of the literature, our properties are combinatorial and finitary in nature, and remain non-trivial over finite structures as well. There has been recent interest in the context of finite structures over finite structures that are reducts of bounded degree, and there has been considerable work on model-theoretic preservation theorems in this context. In particular, recent work by Dawar and Grohe [2] shows that the preservation theorem for classes of structures that are reducts of bounded degree, even when restricted to sentences of bounded logic, fails to hold. This is surprising, since the preservation theorem for classes of structures that are reducts of bounded degree, even when restricted to sentences of bounded logic, holds in general.

In summary, our preservation properties have yielded new preservation theorems in the contexts of both, classical model theory and finite model theory. We have identified in [3] many interesting classes of finite structures that admit the generalized Los-Tarski theorem, and moreover, admit it in effective form. Specific examples include the classes of words, trees (as partial orders), structures of bounded tree-depth, grids of bounded dimension, various well-known subclasses of co-graphs, etc. Some of these, like structures of bounded tree-depth, were earlier not known to even satisfy the Los-Tarski theorem.

Bibliography

