

Graph-based Knowledge Heuristics

Udrea (Dănciulescu) Daniela

Preliminaries. Cybernetics regards information in three aspects: *syntactically*, as a way of representation by numbers, sizes, sounds, etc., *semantically*, from the point of view of meaning for the receiver and *pragmatically*, that is, from the point of view of utility.

Computer technology - bringing together research, design, use and maintenance of automation systems for data processing - tends to become a generalized infrastructure. This will be achieved both by incorporating computer hardware and software into industrial and household products, and especially by the usage of information technologies to assist, regulate and control the technical and productive processes, both creative and leadership.

In the world of information technology there is no clear distinction between "information processing" and "data processing". If, at the beginning, the phrase "Electronic Data Processing" (EDP) was used after the expression "Management Information Systems" (MIS), which is otherwise appropriate because it involves the subjective side given by the presence the human factor, the manager, both in the commercial sphere and in the academic world, the equivalence between information and data has become commonplace. It is true that the term "information" sounds more actual, but let's not forget the semantic effort made by IT to differentiate in the database world the database management systems (DBMS) and data base servers from the software applications and middleware. If in DBMS the term "data" is perfectly suited, in MIS or in the more specialized DSS (Decision Support Systems) and EIS (Executive Information Systems), the term "information" is more justified. Indeed, in the MIS / DSS / EIS area, information is always subjective, while in non-applicative software and middleware information is always objective.

Starting from the informatics system developed during the doctoral thesis: ***Using the Statistical Methods in the Sanitary Information System Improvement***, PhD Supervisor: Prof. Nicolae Vasilescu, Faculty of Economics and Business Administration, University of Craiova, I have studied several aspects related to the implementation of intelligent systems. The SaSome application built during my PhD studies presents the main features of any expert system:

- contains a database and implicitly uses a database management system
- contains an inference engine that has the following basic tasks:
 - it takes inputs from the input interface
 - it processes the information stored in the database
 - it obtains the conclusions generated by processing the data stored in the database
 - follows the execution steps and, depending on the executed steps, formulates the responses to be displayed at the output interface
- in order to facilitate user communication, the application was endowed with a user-friendly graphical interface

The aspects developed in this habilitation thesis and the research directions have the following main objectives:

1. Addressing the theoretical aspects of an IT product quality management. The case study focuses on eLearning products and how they interact with users;
2. The approach of the theoretical aspects in view of the algebraic properties of graphical representations;
3. Hybrid optimization approach based on genetic algorithms. The analyzed case study addresses the issue of vehicle routing;
4. Developing interfaces in natural language:
 - Designing properties of graphs labeled for applications targeting formal languages;
 - Generate natural language using labeled graphs.

The research presented in this thesis is intended for developing new graph-based systems for knowledge representation and also for knowledge reasoning. The main concept, based on which the formalisms presented in the first four chapters of this work are constructed, corresponds to the Labeled Stratified Graphs structure, or shortly, Stratified Graph structure.

The Stratified Graph structure was introduced in the literature by professor Nicolae Țăndăreanu in the year 2000 and is defined by means of a system of components which encode a labeled graph concept: $G = (S, L_0, T_0, f_0)$. Because the labeled graphs are weak structures that can not allow the development of an advanced mechanism of knowledge processing, the stratified graphs formalism makes use of some methods from universal algebra for defining a new knowledge representation and reasoning mechanism.

In **Chapter 1** of this thesis we give the motivation for the studies included in this thesis. Also, we enumerate the involved algebraic concepts. More precisely, in the first chapter we review the algebraic concepts involved in the labeled graph and stratified graph structures: fundamentals of ordered sets such as partial mappings or partial order and partial σ -algebras constructions.

Resuming, this chapter underlines:

- Intelligent systems development technologies: levels of artificial intelligence, calculus and inference;
- Algebraic properties of the representations based on graph structures. Graph-based representations are the basis of the theoretical models presented in Chapters 3 and 4;
- The utility of intelligent systems in the economic field. Chapter 3 is dedicated to the study of the development of a smart system that uses genetic technologies to implement the vehicle problem);
- Natural language processing technologies and their applicability in the economic field. In Chapter 4 there are some studies related to the syntactic nature of language

In **Chapter 2** we present an e-platform project together with its qualitative management criteria. The project was presented in and „supports e-learning methods implementing in the education process focused on project target group, enhancing training attractiveness, providing target group with new ways of gaining knowledge (interactive multimedia presentations, virtual classes, access to the up-to-date project education areas achievements). It also offers flexibility in time and location, cooperation, experience sharing, common European dimension, accessible learning to all target group members, self-directed learning and peer support encouragement, lower training costs (travel/accommodation in academic institutions), language proficiency improvement.

Therefore, the quality assurance assessment criteria were important and focused on the main components of the e-platform. When testing the quality assurance of e-platform and its component modules there were used criteria focused on fitness for purpose, suitability of provided knowledge, user interface, language level, graphic approach and interaction level. Criteria used for assessing software characteristics quality were focused on functionality, reliability, usability, efficiency, maintainability, and portability.

Based on the information gathered from target group regarding module assessment, the following conclusions were provided: the respondents answered questions regarding overall content, ease of navigation and overall look, links, visuals, style as well as content of information provided. The quality of the overall was seen as high to very high by 82 % members of project target group. The quality of the various features of the modules (content and topics of modules, courses, virtual classes etc.) was seen as high to very high by more than 70% of respondents".

The studies presented in this chapter were published in:

1. Ali Amer Mohammed Saeed, Daniela Dănciulescu, *Modern Interfaces for Knowledge Representation and Processing Systems Based on Markup Technologies*, International Journal of Computers Communications & Control (IJCCC), 13(1), pp. 117-128, February 2018, ISSN 1841-9836
2. Gheorghe Grigoraş, Daniela Dănciulescu, Cătălina Sitnikov, *Assessment Criteria of E-learning Environments Quality*, Procedia Economics and Finance, vol 16, pp. 40--46 (2014)
3. Georgeta Şoavă, Cătălina Sitnikov, Daniela Dănciulescu, *Optimizing Quality of a System Based on Intelligent Agents for E-Learning*, Procedia Economics and Finance, vol 16, pp. 47--55 (2014)
4. Gheorghe Grigoraş, Paraschiva Popovici, Daniela Dănciulescu, *The Management and Continuous Improvement Model of Quality in Academic Education from the Perspective of the Education - Reseach Process*, The 3rd International Conference: Institutional Strategic Quality Management (ISQM2011), July 14--16, Sibiu, Romania (2011)
5. Daniela Dănciulescu, Carmen Radu, Iancu Anica, *Formal Theory of the Intelligent Economic Systems*. In: Economic indicators used for EU projects, in other criteria of aggregation than national/regional, pp. 772--775 (2007)

Chapter 3 presents a new approach to the hybrid optimization problem based on genetic algorithms in order to solve the clustered vehicle routing problem (CluVRP). The problem investigated in this chapter, namely the routing problem of clustered vehicles, is a NP-hard

combinatorial optimization problem that generalizes the classic vehicle routing problem (VRP). The latter is directly related to the generalized vehicle routing problem (GVRP) and belongs to the general class of optimization issues. This class generalizes in a natural way the classical combinatorial optimization problems and is characterized by the fact that the nodes are divided into clusters and the constraints are expressed in terms of groups and not individually. The problems I have studied have a strong economic character, being driven by real applications. To solve these problems, we used intelligent algorithmic methods to provide a solution.

The study presented in Chapter 3 follows three directions, each of significant interest in economic research, namely:

- The clustered vehicle routing problem (CluVRP) is studying the determination of an optimal collection of routes or routes to minimize costs for a fleet of vehicles, while respecting the requirements of users driving clustering on the one hand, taking into account the capacity constraints of vehicles and at the same time respecting the additional constraining constraint of all customers belonging to a cluster before leaving it. We proposed a new approach for solving this problem by using a decomposition method that separates the CluVRP problem into two easier subproblems. *The results of the computational experiments on three sets of reference points demonstrate that our approach offers better results than other approaches;*
- The Prize-Collecting Generalized Minimum Spanning Tree Problem (PC-GMSTP) is a variation of the generalized minimum spanning tree problem (GMSTP), where, in addition to the non-negative edge costs, we have non-negative payments for the vertices to be selected. The problem is inspired by a practical application, namely the problem of designing local telecommunication networks. We have devised a new approach to solving the PC-GMSTP problem by decomposing it into two smaller subproblems: a macro subproblem and a micro-level subproblem and solving them separately. The purpose of the first subproblem is to provide cluster coverage trees using a diploid genetic algorithm, and the purpose of the second problem is to determine the minimum cost tree corresponding to previous global trees, covering exactly one vertex for each cluster. The second problem is effectively resolved using dynamic programming. Preliminary computational results obtained using a standard set of benchmarks demonstrate that our approach offers better solutions than the other existing PCS-GMSTP problem solving methods;
- We presented a new solution for structuring information in a relational database system. Our approach uses concepts inspired from object oriented programming and differs from other studies to solve the same problem. The new solution found by us combines library information standards with object oriented programming techniques to manage the financial and contact information of potential clients and partners.

Given a depot and a set of customers divided in a number of clusters, the problem aims at finding an optimal collection of tours (w.r.t. cost minimization) from depot to the customers satisfying the capacity constraints and the supplementary restriction that all the customers pertaining to a

particular cluster are visited consecutively before visiting some other customers. CluVRP was introduced by Sevaux and Sorensen and is directly related to the following issues:

- The Generalized Vehicle Routing Problem (GVRP) consists in designing the optimal distribution of collection routes from the depot to customers, routes that are subject to the capacity constraints and additional constraints due to the fact that each group should be visited exactly as a node. GVRP was first introduced by Ghiani and Improta.
- The issue of optimal distribution of collection routes from the depot to customers subject to capacity constraints but also to additional constraints due to the fact that at least one node must be visited from each group. This problem was introduced in the literature by Baldacci and Laporte.

We have defined two kinds of edges: the edges of nodes belonging to the same cluster, called intra-cluster edges, and the edges of nodes belonging to different clusters, called inter-cluster edges. Graph G is supposed to be strongly connected and generally supposed to be complete.

The clustered vehicle routing problem (CluVRP) scope is to find a collection of routes in order to visit all clusters and all low cost nodes so that the following constraints occur: each route starts and ends at the top of the depot, all the nodes in each cluster must be visited consecutively and the sum of requests of the visited nodes by a route does not exceed the capacity of the vehicle, Q .

The results of the preliminary calculations show that our hybrid algorithm behaves favorably in terms of solution quality compared to existing approaches. In the future, we aim to check the generality and scalability of the proposed hybrid heuristic approach by testing it across multiple instances and combine them with local search procedures to refine the solutions provided by our algorithm.

The studies presented in this chapter were published in:

1. Petrică C. Pop, Oliviu Matei, Corina Pop Sitar, Daniela Dănciulescu, *Genetic Algorithm Based Solution Approach to Solve the Prize-Collecting Generalized Minimum Spanning Tree Problem*, CIE47 Proceedings, 11-13 October 2017, Lisbon Portugal (accepted 2017)
2. Andrei Horvat Marc, Levente Fuksz, Petrică C. Pop, Daniela Dănciulescu (2016) *A decomposition-based method for solving the Clustered Vehicle Routing Problem*, Journal of Applied Logic, ISSN: 1367—0751
3. Cosmin Sabo, Petrică C. Pop, Honoriu Vălean, Daniela Dănciulescu, *An Innovative Approach to Manage Heterogeneous Information Using Relational Database Systems*, Proceedings of International Conference on Intelligent Systems Design and Applications. ISDA 2016: Intelligent Systems Design and Applications, pp 1-10 (2016)
4. Marc, A. H., Fuksz, L, Pop, P. C., Dănciulescu, D. (2015) *A Novel Hybrid Algorithm for Solving the Clustered Vehicle Routing Problem*, Proceedings of International Conference HAIS 2015, Springer International Publishing Switzerland
5. Amelia Badică, Costin Badică, Florin Leon, Daniela Dănciulescu, *Optimization of Freight Transportation Brokerage Using Agents and Constraints*. Communications in Computer and

Chapter 4 is dedicated to the defining of a new mechanism for generating formal languages. This mechanism, also based on stratified graph representations, can generate first and second type languages. Specifically, this chapter proposes a new system for formal language generation through a Stratified Graph Representation System. It is proved that by using a dedicated interpretation system for this representation system, a particular formal language can be obtained through structured accepted paths from a stratified graph.

There are several mechanisms for generating regular languages in the literature, such as grammars, automata, transition networks, etc. In Chapter 4, we consider the case of the generation of Right Linear Languages, and we demonstrate that a Stratified Graph Representation System can model a right-linear grammar by means of an appropriate interpretation system.

In a future study, we intend to investigate the way in which, using a set of restrictions in the inference process, the sequences of the generated formal language will be (or are not) affected. We also intend to study families of formal languages that can be obtained through this representation system; for example, we want to test the ability to generate regular or context-sensitive languages.

Natural language processing involves a set of computational tools for analyzing and extracting conclusions from the human (natural) text. Different features of natural documents (words, phrases, emoticons, etc.) are identified and new information is generated on human like classifications and assessments, individual decisions, etc. In a corpus, a document can be represented as a vector of features with associated meanings. The likelihood of expressing a feeling in a document or approaching a particular subject can be evaluated using these vectors. Many search engines use autocorrect algorithms in which this type of vector is implemented. The main reasons for using NLP tools in many areas of activity are:

- The possibility of extending the available data sources for the study. NLP can create a basis for econometric studies based on information extracted from user reviews, employee ratings, news, and other sources of text. Online documents and behavior within social networks are excellent sources that can be explored at minimal cost compared to other traditional methods such as surveys.
- The ability to anticipate unobserved results. Reviews and other texts that contain explanatory or forward-looking information can be used to perform what-if analyzes or to generate scenarios useful in assisting decision-making processes.
- The ability to understand the factors that lead to certain results. In NLP models, forward-looking information is associated with words or parts of a sentence with a certain meaning and gives an idea of the type of feelings, information, emotions that lead to certain results.

At present, more and more NLP tools are used in the economic field. These tools can help economists in various industries:

- **Marketing.** NLP tools are very useful in the field of advertising, branding and customer relationship management. Brand awareness campaigns and brand awareness campaigns can be developed based on customer sentiment analysis. For Customer Relationship Management (CRM), interactions between customers and companies are a useful indicator of discovering the reasons for customer dissatisfaction and intentions. Recently, the company that offers Sugar CRM CRM has acquired Contastics NLP technology, a platform that analyzes emails, LinkedIn and other sources from sellers and customers.
- **Financial services.** Documents, textiles, comments, reviews are sources of unstructured information that can be combined with other traditional financial analysis tools, providing much more accurate information about financial markets.
- **Audit.** NLP tools can enrich traditional databases used for accounting audit by using mining text techniques on unofficial sources.
- **Human capital management.** Recruiting new or highly trained staff requires high costs for businesses. NLP techniques can assist these activities, leading to lower costs.
- **Reputation monitoring.** To survive in a global environment, companies need tools that allow them to analyze external data, public data, or open-source data. Some of these data are unstructured and are derived from information sharing media such as blogs, news sites, forums, online reports, social networks, business sites, and more. Many customers check online reviews before purchasing a product or contacting a company for a service, so monitoring and managing reputation is a priority for businesses. They can scan for web-related items or company names and identify situations that need to be fixed or requiring action.

The studies presented in this chapter were published in:

1. Dănciulescu, D. (2015) *Formal Languages Generation in Systems of Knowledge Representation based on Stratified Graphs*, INFORMATICA, vol. 26, no. 3, pp. 407-417, ISSN 0868-4952.
2. Daniela Dănciulescu, Mihaela Colhon, Gheorghe Grigoraș, *A System of Knowledge Representation for Right Linear Grammars Generation*, Broad Research in Artificial Intelligence and Neuroscience, vol 8(1), ISSN 2068-0473, E-ISSN 2067-3957, pp. 42-51 (2017)
3. Viorel Negru, Gheorghe Grigoraș, Daniela Dănciulescu, *Natural Language Agreement in the Generation Mechanism based on Stratified Graphs*, Proceedings of the 7th Balkan Conference in Informatics (BCI 2015), Craiova, Romania, pp. 36:1--36:8, DOI: 10.1145/2801081.2801121 (2015)
4. Daniela Dănciulescu, Mihaela Colhon, *Systems of knowledge representation based on stratified graphs. Application to Natural Language Generation*, Carpathian Journal of Mathematics, 32(1), pp. 49--62 (2014)

5. Daniela Dănciulescu, Mihaela Colhon, *Splitting the structured paths in stratified graphs. Application in Natural Language Generation*, Analele Științifice ale Universității Ovidius Constanța, Seria Matematică , vol. 22, no. 2, pp.59-69, ISSN: 1224-1784 (2014)
6. Daniela Dănciulescu, Nicolae Țândăreanu, *Splitting the structured paths in stratified graphs, Modeling and Development of Intelligent Systems*, Proceedings of the 3th Int Conference on Modeling and Development of Intelligent Systems, Sibiu 10-12 Oct. 2013, University of Lucian Blaga from Sibiu Publishing House, ISSN 2067-3965 (2014)
7. Daniela Dănciulescu, *Systems Of Knowledge Representation Based On Stratified Graphs And Their Inference Process*, 9th International Conference of Applied Mathematics, Abstracts and Pre-Proceedings, Baia Mare 25-28 September (2013)
8. Mihaela Colhon, Dana Dănciulescu, *Semantic Schemas for Natural Language Generation in Multilingual Systems*, Journal of Knowledge, Communications and Computing Technologies (JKCCT), ISSN: 2067-0958, vol II(1)/2010, pp. 10-17 (2010)

Conclusions and Future Directions

At the University of Craiova, the Department of Computer Science, there is a research center (RCAI, <http://www.rcai.eu/>), which has an experienced research group active in the field of Artificial Intelligence and the applications of this research field in e-learning, economics and communications. Being part of this research group as the director of the group I intend to develop its research team by writing research projects and organizing round tables with researchers in the field and with companies that activate in this scientific domain, in order to identify useful applications for both parties.