

Graph-based Knowledge Heuristics

(Cunoștințe euristice cu reprezentări de tip graf)

Udrea (Dănciulescu) Daniela

Preliminarii. Cibernetica privește informația sub trei aspecte: *sintactic*, ca mod de reprezentare prin numere, mărimi, sunete etc, *semantic*, din punct de vedere al sensului (semnificației) pentru cel ce o recepționează și *pragmatic*, adică din punct de vedere al utilității.

Tehnologia informatică - reunind cercetarea, proiectarea, folosirea și întreținerea sistemelor de automatizare a prelucrării datelor - tinde să devină o infrastructură generalizată. Aceasta se va realiza atât prin încorporarea de elemente informatice - hardware și software - în produse industriale și de uz casnic, cât mai ales prin folosirea tehnologiilor informatice în asistarea, reglarea și controlul proceselor tehnico-productive, a celor creative și de conducere.

În lumea tehnologiei informației nu apare o distincție clară între “prelucrarea informației” și “prelucrarea datelor”. Dacă la început se folosea sintagma «Prelucrarea electronică a datelor» - Electronic Data Processing (EDP), după apariția sintagmei “Sisteme de gestionare a informației” - Management Information Systems (MIS), care este de altfel adecvată deoarece implică latura subiectivă dată de prezența factorului uman, managerul, atât în sfera comercială cât și în lumea academică echivalența între informație și date a devenit obișnuită. E adevărat că termenul «informație» sună mai actual, dar să nu uităm de efortul semantic făcut de informaticieni de a diferenția, în lumea bazelor de date, softul de gestiune - data base management systems (DBMS) și data base servers, de aplicații software și middleware. Dacă în DBMS “data” este perfect adecvat, în domeniul MIS (și al mai specializateelor DSS (Decision Support Systems) și EIS (Executive Information Systems)), termenul “informație” este mai justificat. Într-adevăr, în zona MIS/DSS/EIS, informația este întotdeauna subiectivă, în timp ce în software-ul non-aplicativ și în middleware informația este întotdeauna obiectivă.

Plecând de la sistemul informatic dezvoltat în timpul tezei de doctorat: **Utilizarea Metodelor Statistice în Perfectionarea Sistemului Informatic din Domeniul Sanitar**, Conducator Doctorat: Prof. Nicolae Vasilescu, Facultatea de Economie și Administrarea Afacerilor, Universitatea din Craiova, am studiat mai multe aspecte ce țin de implementarea sistemelor inteligente. Aplicația **SaSome** construită în perioada studiilor doctorale prezintă caracteristicile principale ale unui sistem expert:

- conține o bază de date și implicit utilizează un sistem de gestiune de baze de date
- conține un motor de inferențe care are următoarele sarcini de bază:
 - preia intrările din interfața de input
 - prelucrează informațiile stocate în baza de date
 - obține concluziile generate de prelucrările generate și de datele stocate în bază
 - urmărește etapele de execuție și în funcție de etapele executate formulează răspunsurile ce vor fi afișate la interfața de output
- pentru a facilita comunicarea cu utilizatorul, aplicația a fost înzestrată cu o interfață grafică ușor de manipulat de către utilizator

Aspectele dezvoltate în teza de abilitare și direcțiile de cercetare au următoarele obiective principale:

1. Abordarea aspectelor teoretice ce țin de managementul calității produselor informatice. Studiul de caz vizează produsele de e-Learning și modul în care acestea interacționează cu utilizatorii;
2. Abordarea aspectelor teoretice în viziunea proprietăților algebrice ale reprezentărilor bazate pe grafuri;

3. Abordare hibridă de optimizare bazată pe algoritmi genetici. Studiul de caz analizat vizează problema rutării vehiculelor;
4. Dezvoltarea de interfețe în limbaj natural:
 - a. Proiectarea proprietatilor grafurilor etichetate la aplicatii ce vizeaza limbajele formale;
 - b. Generarea de limbaj natural cu ajutorul grafurilor etichetate.

Rezumatul tezei pe capitole:

Capitolul 1 conține motivarea studiilor teoretice prezentate în teză:

- Sistemele inteligente: nivele de inteligență artificială, calculabilitate și inferență;
- Proprietățile algebrice ale reprezentărilor bazate pe structuri de tip graf. **Reprezentările de tip graf stau la baza modelelor teoretice prezentate în Capitolele 3 și 4;**
- *Utilitatea sistemelor inteligente în domeniul economic. Capitolul 3 este dedicat studiului dezvoltării unui sistem inteligent ce folosește tehnologii genetice în vederea implementării problemei vehiculelor);*
- Tehnologii de procesare a limbajului natural și aplicabilitatea acestora în domeniul economic. În Capitolul 4 sunt prezentate câteva studii legate de natura sintactică ale limbajului.

Capitolul 2 este dedicat criteriilor de managementul calității pentru sistemele e-learning. Așa cum se arată în literatura de specialitate, o platformă de e-learning sprijină implementarea metodelor de învățare electronică în procesul de învățământ axat pe utilizator, sporirea atractivităților de formare, oferirea de noi modalități de dobândire a cunoștințelor (prezentări multimedia interactive, cursuri virtuale etc.). De asemenea, o platformă de e-learning oferă flexibilitate în timp și locație, cooperare, schimb de experiență, o dimensiune europeană comună, învățare accesibilă tuturor utilizatorilor, încurajarea auto-dirijată și încurajarea sprijinului reciproc, costuri reduse de instruire (călătorii / cazare în instituții academice).

Prin urmare, criteriile de evaluare a asigurării calității au fost importante și s-au axat pe componentele principale ale platformei electronice. La testarea asigurării calității platformei electronice și a modulelor sale componente au fost utilizate criterii axate pe aptitudinea de a fi folosite, cunoștințele furnizate, **interfața cu utilizatorul**, nivelul de limbă, abordarea grafică și nivelul de interacțiune. Criteriile utilizate pentru evaluarea calității caracteristicilor software s-au axat pe funcționalitate, fiabilitate, utilitate, eficiență, întreținere și portabilitate.

Studiile prezentate în acest capitol au fost publicate în:

1. Ali Amer Mohammed Saeed, Daniela Dănculescu, *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 (IF=1.374)
2. Gheorghe Grigoraș, Daniela Dănculescu, Cătălina Sitnikov, *Assessment Criteria of E-learning Environments Quality*, Procedia Economics and Finance, vol 16, pp. 40--46 (2014)
3. Georgeta Șoavă, Cătălina Sitnikov, Daniela Dănculescu, *Optimizing Quality of a System Based on Intelligent Agents for E-Learning*, Procedia Economics and Finance, vol 16, pp. 47--55 (2014)
4. Gheorghe Grigoraș, Paraschiva Popovici, Daniela Dănculescu, *The Management and Continuous Improvement Model of Quality in Academic Education from the Perspective of the Education - Research 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)

Capitolul 3 prezintă o nouă abordare a problemei de optimizare hibridă bazată pe algoritmi genetici pentru rezolvarea problemei de rutare a vehiculelor grupate (CluVRP). Problema investigată în acest capitol, și anume problema de rutare a vehiculelor grupate este o problemă de optimizare combinatorică NP-hard care generalizează problema clasică de rutare a vehiculelor (VRP). Aceasta din urmă este direct legată de problema generalizată de rutare a vehiculelor (GVRP) și aparține clasei generale de probleme de optimizare. Această clasă generalizează într-un mod natural problemele clasice de optimizare combinatorială și este caracterizată de faptul că nodurile sunt partiționate în clustere, iar constrângerile sunt exprimate în termeni de grupuri și nu individual. Problemele pe care le-am studiat au un puternic caracter economic, fiind izvorate din aplicații reale. Pentru rezolvarea acestor probleme am folosit metode algoritmice inteligente pentru a furniza o soluție.

Studiul prezentat în capitolul 3 urmărește trei direcții, fiecare de interes semnificativ în domeniul cercetării economice și anume:

- Problema rutării vehiculelor grupate (CluVRP) studiază determinarea unei colecții optime de rute sau trasee în vederea minimizării costurilor, pentru o flotă de autovehicule, cu respectarea cerințelor utilizatorilor care conduc la gruparea vehiculelor în clustere pe de o parte, având în vedere constrângerile de capacitate a vehiculelor și totodată respectând constrângerea suplimentară de vizitare consecutivă a tuturor clienților aparținând unui cluster înainte de părăsirea acestuia. Am propus o nouă abordare pentru rezolvarea acestei probleme prin utilizarea unei metode bazate pe decompoziție care separă problema CluVRP în două subprobleme mai ușor de rezolvat. *Rezultatele experimentelor computaționale asupra a trei seturi de repere de referință din literatură demonstrează că abordarea noastră oferă rezultate mai bune comparativ cu alte abordări;*
- Problema generalizată a arborelui de acoperire de cost minim cu colectare de premii (PC - GMSTP) este o variantă a problemei generalizate a arborelui de acoperire de cost minim (GMSTP), unde, în plus față de costurile nenegative atașate muchiilor, avem plăți nenegative pentru vârfurile care se selectează. Problema este inspirată dintr-o aplicație practică și anume problema proiectării rețelelor de telecomunicații locale. Am conceput o nouă abordare pentru rezolvarea problemei PC-GMSTP prin descompunerea acesteia în două subprobleme mai mici: o subproblemă la nivel macro și o subproblemă la nivel micro și prin rezolvarea separate a acestora. Scopul primei subprobleme este de a furniza arbori de acoperire a clusterelor folosind un algoritm genetic diploid, iar scopul celei de-a doua probleme este de a determina arborele de cost minim corespunzător arborilor globali anteriori, acoperind exact un vârf pentru fiecare cluster. Cea de-a doua problemă se rezolvă eficient folosind programarea dinamică. Rezultatele computaționale preliminare obținute folosind un set usual de repere de referință demonstrează că abordarea noastră oferă soluții mai bune comparativ cu alte metode existente pentru rezolvarea problemei PC-GMSTP;
- Am prezentat o nouă soluție pentru structurarea informației într-un sistem de baze de date relaționale. Abordarea noastră utilizează concepte inspirate din programarea orientată obiect și diferă de alte studii pentru rezolvarea aceleiași probleme. Soluția novativă găsită de noi combină standardele informaționale din domeniul bibliotecilor cu tehnicile de programare orientată obiect, în vederea gestionării informațiilor financiare și de contact ale potențialilor clienți și parteneri.

Fiind dat un depozit și o mulțime de clienți care sunt grupați într-un număr de grupuri predefinite, CluVRP constă în găsirea optimă a distribuiri sau a rutelor de la depozit la clienți care respectă constrângerile de capacitate, dar și constrângerile generate de faptul că toți clienții care aparțin unui grup trebuie să fie vizitați consecutiv. CluVRP a fost introdusă de Sevaux și Sorensen și este direct legată de următoarele probleme:

- problema generalizată de rutare a vehiculelor (GVRP) care constă în proiectarea distribuiri optime a rutelor de colectare din depozit către clienți, rute care sunt supuse unor constrângeri de capacitate dar și unor constrângeri adiționale datorate faptului că fiecare grup trebuie vizitat exact ca un nod. GVRP a fost introdus pentru prima dată de Ghiani și Imbrota.

- problema distribuiri optime a rutelor de colectare din depozit către clienți supuse unor constrângeri de capacitate dar și unor constrângeri adiționale datorate faptului că din fiecare grup trebuie vizitat cel puțin un nod. Această problemă a fost introdusă în literatura de specialitate de către Baldacci și Laporte.

Am definit două feluri de muchii: muchii dintre vârfuri care aparțin aceluiași cluster, numite muchii intra-cluster, și muchii dintre vârfuri care aparțin unor cluster diferite, numite muchii inter-cluster. Graful G se presupune că este puternic conectat și, în general, se presupune că este complet.

Problema de rutare a vehiculelor grupate (CluVRP) constă în găsirea unei colecții de rute care să viziteze toate clusterurile și toate vârfurile cu cost minim, astfel încât următoarele constrângeri au loc: fiecare rută începe și se termină la vârful depozit, toate vârfurile din fiecare cluster trebuie să fie vizitate consecutiv și suma cererilor vârfurilor vizitate de o rută să nu depășească capacitatea vehiculului, Q .

Rezultatele calculelor preliminare arată că algoritmul nostru hibrid se comportă favorabil în termeni de calitate a soluției în comparație cu abordările existente. Pe viitor, ne propunem să verificăm generalitatea și scalabilitatea abordării euristice hibride propuse prin testarea ei pe mai multe instanțe și să le combinăm cu proceduri de căutare locale pentru a rafina soluțiile oferite de algoritmul nostru.

Studiile prezentate în acest capitol au fost publicate în:

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 Information Science Volume 744, 2017, Pages 451-464, 18th International Conference on Engineering Applications of Neural Networks, EANN 2017; Athens; Greece

Capitolul 4 este dedicat definirii unui nou mecanism de generare de limbaje formale. Acest mecanism, bazat de asemenea pe reprezentări de tip graf stratificat, poate genera limbaje de primul și de al doilea tip. Mai exact, în acest capitol se propune un nou sistem pentru generarea de limbaj formal prin intermediul unui Sistem de Reprezentare bazat pe Grafuri Stratificate. Se validează că, prin folosirea unui

sistem de interpretare dedicat pentru acest sistem de reprezentare, un limbaj formal particular poate fi obținut prin intermediul drumurilor structurate acceptate dintr-un graf stratificat.

În literatura de specialitate există mai multe mecanisme de generare de limbaje regulate, ca de exemplu: gramatici, automate, rețele de tranziție, etc. În Capitolul 4 considerăm cazul generării Limbajelor Liniare la Dreapta și demonstrăm ca un Sistem de Reprezentare bazat pe Grafuri Stratificate poate sa modeleze o gramatică liniară la dreapta prin intermediul unui sistem de interpretare adecvat.

Într-un studiu viitor intenționăm să investigăm modalitatea prin care, folosind un set de restricții în calculele procesului de inferență, secvențele limbajului formal generat vor fi (sau nu) afectate. De asemenea, intenționăm să studiem familiile de limbaje formale care pot fi obținute prin acest sistem de reprezentare; de exemplu, dorim să testăm posibilitatea de generare a limbajelor regulate sau a limbajelor context-sensitive.

Prelucrarea limbajului natural cuprinde o mulțime de unelte computaționale pentru analiza și extragerea de concluzii din textul uman (natural). Se identifică diverse caracteristici ale documentelor natural (cuvinte, fraze, emoticoane etc.) și se creează informație nouă cu privire la clasificările și evaluările umane, deciziile individuale etc. Într-un corpus, un document se poate reprezenta sub forma unui vector de caracteristici cu semnificații asociate. Probabilitatea de exprimare a unui sentiment într-un document sau de abordare a unui anumit subiect pot fi evaluate folosind acești vectori. Foarte multe motoare de căutare și algoritmi de corecție (autocorrect) utilizează acest tip de vectori. Principalele motive pentru utilizarea uneltelor NLP în toate domeniile de activitate sunt:

- Posibilitatea de extindere a surselor de date disponibile pentru studiu. NLP poate crea o bază pentru studii economice pe baza informațiilor extrase din recenziile utilizatorilor, evaluările angajaților, știri și alte surse de texte. Documentele online și comportamentul în cadrul rețelelor sociale reprezintă surse excelente care pot fi explorate cu costuri minime comparative cu alte metode tradiționale cum ar fi sondajele.
- Posibilitatea de a anticipa rezultatele neobservate. Recenziile și alte texte care conțin informații explicative sau anticipative pot fi utilizate pentru realizarea de analize de tip what-if sau pentru generarea de scenarii utile în asistarea proceselor de luare a deciziilor.
- Posibilitatea de a înțelege factorii care conduc la anumite rezultate. În cadrul modelelor NLP, informațiile anticipative sunt asociate unor cuvinte sau părți de propoziție cu un anumit înțeles și oferă o idee asupra tipului de sentimente, informații, emoticoane care conduc la anumite rezultate.

La ora actuală, din ce în ce mai multe unelte de tip NLP sunt utilizate în domeniul economic. Aceste unelte pot ajuta economiștii în diverse ramuri de activitate:

- **Marketing.** Uneltele NLP sunt foarte utile în domeniul reclamei, al creării mărcilor și al managementului relațiilor cu clienții. Se pot dezvolta campanii de promovare și de conștientizarea mărcii (brandului) pe baza analizei sentimentelor clienților. Pentru managementul relațiilor cu clienții (CRM), interacțiunile dintre clienții și companii reprezintă o sursă de indicii utile în descoperirea motivelor insatisfacției clienților și a intențiilor acestora. De curând, compania care oferă CRM-ul Sugar CRM a achiziționat Contastics NLP technology, o platformă care analizează comunicările din emailuri, LinkedIn și alte surse dintre vânzători și clienți.
- **Servicii financiare.** Documentele, textele, comentariile, recenziile reprezintă surse de informații nestructurate ce pot fi combinate cu alte unelte tradiționale de analiză financiară, oferind informații mult mai precise despre piețele financiare.
- **Audit.** Uneltele NLP pot îmbogăți bazele de date tradiționale utilizate pentru auditul contabil, prin folosirea tehnicilor de text mining pe surse neoficiale.

- **Managementul capitalului uman.** Recrutarea de personal nou sau foarte bine pregătit profesional necesită costuri ridicate pentru firme. Tehnicile NLP pot asista aceste activități, ducând la scăderea costurilor.
- **Monitorizarea reputației.** Pentru a supraviețui într-un mediu global, companiile au nevoie de unelte care să le permită să analizeze date externe, date publice sau open-source. Unele dintre aceste date sunt nestructurate și sunt obținute din medii de partajare a informației ca blogurile, site-urile de știri, forumurile, rapoartele online, rețelele sociale, site-urile firmelor etc. Foarte mulți clienți verifică recenziile online înainte de a achiziționa un produs sau a contacta o firmă pentru un serviciu, prin urmare monitorizarea și managementul reputației reprezintă o prioritate pentru firme. Acestea pot scana postările web legate de produsele sau numele companiei și identifica situațiile care trebuie remediate sau care necesită anumite acțiuni.

Studiile prezentate în acest capitol au fost publicate în:

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)

Concluzii și Direcții viitoare

La Universitatea din Craiova, la Departamentul de Informatică, există un centru de cercetare (RCAI; <http://www.rcai.eu/>), care are un grup de cercetare experimentat, ce activează în domeniul Inteligenței Artificiale și aplicațiile acestui domeniu de cercetare în e-learning, economie și comunicații. Din acest grup de cercetare fac parte și eu, în calitate de director și intenționez să dezvolt acest grup, prin scrierea de proiecte de cercetare și organizarea de mese rotunde atât cu cercetători în domeniu, cât și cu firme care au în domeniul de activitate cercetarea, în vederea identificării unor aplicații utile pentru ambele părți.