

Departamentul de Informatică
Universitatea din Craiova, România

Mihaela Colhon

Baze de cunoștințe

Note de curs

2014

Cuprins

1	Aspecte de reprezentare a cunoștințelor	11
1.1	Tipuri de cunoștințe	11
1.1.1	Metode pentru reprezentarea cunoștințelor	13
1.1.2	Cunoștințe relaționale simple	15
1.1.3	Cunoștințe care permit moștenirea	15
1.1.4	Cunoștințe inferențiale	16
1.1.5	Cunoștințe procedurale	16
1.2	Sisteme de Reprezentare și Procesare	17
1.2.1	Sisteme bazate pe cunoștințe	18
1.2.2	Motor de inferență	20
1.2.3	Reprezentarea și procesarea cunoștințelor în Prolog . .	21
1.2.4	Sisteme Expert	24
2	Reprezentarea cunoștințelor bazată pe cadre	29
2.1	Baze de cunoștințe cu frameuri	30
2.2	Inferența în sistemele cu frameuri	34
2.2.1	Sisteme cu frameuri. O singură moștenire	34
2.2.2	Sisteme cu frameuri. Moșteniri multiple	34
2.2.3	Funcția de răspuns	35
3	Sisteme Lindenmayer	41
3.1	Sisteme de rescriere	41
3.1.1	Grafica Turtle	45
3.1.2	Rescrierea arcelor și rescrierea nodurilor	46
3.1.3	Sisteme Lindenmayer cu paranteze	50
3.2	Trăsătura de autosimilaritate. Fractali	51
3.3	Curbe FASS	53
3.4	Implementarea în Prolog a sistemelor DOL	57

4	Reprezentarea și procesarea cunoștințelor în sisteme de producție	59
4.1	Reguli if-then	59
4.2	Motorul de inferență	60
4.2.1	Formalizarea raționamentului înainte	62
4.2.2	Formalizarea raționamentului înapoi	64
4.2.3	Strategii de control	65
4.3	Arbori de decizie	65
4.3.1	Interpretarea fotografiilor spațiale	66
4.4	Modulul de generare a explicațiilor	69
4.5	Incertitudine în sistemele de reguli	69
4.5.1	Factori de certitudine. Întărirea încrederii	72
4.6	Sisteme expert bazate pe reguli	72
5	Mulțimi Rough	75
5.1	Indiscernabilitate	76
5.2	Apartenența nuanțată	78
5.3	Aplicații ale mulțimilor rough	79
6	Reprezentarea cunoștințelor în structuri de tip graf	83
6.1	Grafuri conceptuale	83
6.2	Rețelele semantice	85
6.2.1	WordNet	93
6.2.1.1	Relații semantice în WordNet	94
6.3	Grafuri stratificate etichetate	96
6.3.1	Grafuri stratificate etichetate în sintaza de imagini	99
6.4	Scheme semantice	101
7	Reprezentarea cunoștințelor în Procesarea Limbajului Natural	105
7.1	Clase de gramatici și limbaje	108
7.1.1	Reprezentarea gramaticilor în Prolog	110
7.2	Reprezentări bazate pe structuri de tip graf	114
7.2.1	Rețele de tranziție	115
7.2.2	Traducătoare finite	118
7.3	Analiza sintactică bazată pe constituenți	121
7.4	Analiza sintactică cu relații de dependență	126
7.5	Dezambiguizarea automată a sensului cuvintelor	135
7.5.1	Dezambiguizarea bazată pe cunoștințe	138

Cuvânt înainte

Prezentările din această carte au la bază studiile pe care le-am realizat în cei 15 ani care s-au scurs decând am devenit cadru didactic al Facultății de Matematică și Informatică (actualmente, Facultatea de Matematică și Științe ale Naturii) și până acum. Activitatea mea de cercetare s-a axat pe dezvoltarea de noi mijlocuri de reprezentare și procesare a cunoștințelor. De asemenea, domeniul în care mi-am susținut teza de doctorat cât și tema de cercetare abordată se încadrează în acest domeniul de cercetare, titlul tezei mele de doctorat fiind *Metode de tip graf pentru reprezentarea și procesarea cunoștințelor. Formalisme și reprezentări*. În ultimii 5 ani accentul s-a pus analiza de texte în limbaj natural.

Tot de 15 ani predau *Baze de cunoștințe* studenților: până în 2011 am predat seminariile acestei discipline (titular de curs fiind Prof. dr. Nicolae Țândăreanu) iar din 2011 predau și cursul. Tematica actuală abordată la curs include prezentarea detaliată a celor mai importante sisteme de reprezentare a cunoștințelor: reprezentarea și procesarea cunoștințelor în sisteme cu reguli de producție, reprezentarea și procesarea cunoștințelor structurate (sisteme cu cadre, rețele semantice), sisteme de raționament incert, reprezentarea cunoștințelor în procesarea limbajului natural. În cadrul cursurilor ce tratează reprezentarea prin rețele semantice, am inclus două formalisme derivate din aceste reprezentări: Grafuri Statificate Etichetate și Scheme Semantice dar și prezentarea unui exemplu de rețea semantică extrem de folosită în aplicațiile inteligenței artificiale în general, cu precădere în subdomeniul procesării limbajului natural. Este vorba de rețeaua semantică *WordNet*, datorită căreia procesarea cunoștințelor a dobândit noi dimensiuni. Dincolo de aspectele teoretice ale conceptelor enumerate anterior, în timpul seminariilor am urmărit cu precădere tratarea aplicațiilor practice, căutând să perfecționez abilitățile de programare ale studenților pentru realizarea și utilizarea eficientă a sistemelor de raționament, propunându-le spre analiză implementări variate în diverse limbaje de programare cum ar fi Prolog, Lisp sau Java.

Această carte sintetizează activitatea mea de cercetare în domeniul reprezentării și procesării cunoștințelor precum și materialele didactice și aplicațiile

dezbătute cu studenții în timpul orelor de curs și seminar/laborator.

Deoarece această carte se adresează în primul rând studenților interesați de cursul de *Baze de cunoștințe*, fundamentele teoretice și detaliile de formalizare și implementare ale modelelor de reprezentare pe care le-am inclus în carte și care sunt rodul activităților mele de cercetare sunt tratate pe scurt. Pentru cititorul interesat listez mai jos lucrările în cauză care mi-au servit la redactarea acestei cărți.

Reprezentări de cunoștințe în limbaj natural

Bădică, C., Colhon, M., Șendre, A. *Sentiment Analysis Of Tourist Reviews: Data Preparation And Preliminary Results* In Mihaela Colhon, Adrian Iftene, Verginica Barbu Mititelu, Dan Cristea, Dan Tufiş (eds.) Proceedings of the 10th International Conference "Linguistic Resources And Tools For Processing The Romanian Language", pp. 135-142 (2014)

Colhon, M., Șendre, A., Bădică, C. *Relating the Opinion Holder and the Review Accuracy in Sentiment Analysis of Tourist Reviews*, Proceedings of The 7th International Conference on Knowledge Science, Engineering and Management, KSEM 2014 (2014)

Bibiri, A., Colhon, M., Diac, P., Cristea, D. *Statistics Over A Corpus Of Semantic Links: "QuoVadis"*. In Mihaela Colhon, Adrian Iftene, Verginica Barbu Mititelu, Dan Cristea, Dan Tufiş (eds.) Proceedings of the 10th International Conference "Linguistic Resources And Tools For Processing The Romanian Language", pp. 33-44 (2014)

Cristea, D., Gifu, D., Colhon, M., Diac, P., Bibiri, A.-D., Măranduc, Scutelnicu, L.-A. *Quo Vadis: A Corpus of Entities and Relations*. In N.Gala, R.Rapp and G.B.Enguix (eds.): Language Production, Cognition, and the Lexicon, Springer International Publishing Switzerland (2014)

Colhon, M., Diac, P., Măranduc, Perez, A. "QuoVadis Research Areas - Text Analysis". In Mihaela Colhon, Adrian Iftene, Verginica Barbu Mititelu, Dan Cristea, Dan Tufiş (eds.) Proceedings of the 10th International Conference "Linguistic Resources And Tools For Processing The Romanian Language", pp. 45-56 (2014)

Dănciulescu, D., Colhon, M. *Splitting the structured paths in stratified graphs. Application in Natural Language Generation*, An. St. Univ. Ovidius Constanța, 22(2) pp. 59-69 (2014)

Colhon, M. *Automatic Lexical Alignment Between Syntactically Weak Related Languages. Application for English and Romanian*, Proceedings of ICCCI 2013, vol. 8083, Lecture Notes in Computer Science series (2013)

Hristea, F., Colhon, M. *Feeding Syntactic Versus Semantic Knowledge to a Knowledge-lean Unsupervised Word Sense Disambiguation Algorithm with an Underlying Naive Bayes Model*, Fundamenta Informaticae Journal, ISSN:

- 0169-2968 vol. 119(1), pp. 61-86, DOI 10.3233/FI-2012-727 (2012)
- Colhon, M. *Acquiring Syntactic Translation rules from a Parallel Treebank*, Journal of Information and Library Science INFOtheca, vol. XIII, pp. 19-32, Serbian Academic Library Association, ISSN: 1450-9687 (2012)
- Colhon, M., Simionescu, R. *Deriving a Statistical Syntactic Parsing from a Treebank*, Proceedings of International Conference on Web Intelligence, Mining and Semantics WIMS' 12, ISBN: 978-1-4503-0915-8 (2012)
- Colhon, M. *Language Engineering for Syntactic Knowledge Transfer*, Computer Science and Information Systems Journal (ComSIS), 9(3), ISSN 1820-0214, pp. 1231-1247 (2012)
- Colhon, M. *Syntactic Translation Patterns from a Parallel Treebank*, Proceedings of the First Workshop on Computational Linguistics of Balkan Languages (CLOBL 2012), The 5th Balkan Conference in Informatics - BCI 2012, ISBN 978-86-7031-200-5, pp. 85-88 (2012)
- Colhon, M. *An Architecture Model for English-Romanian Noun Phrase Alignments*, In JKCCT (Journal of Knowledge, Communications and Computing Technologies), ISSN: 2067-095, 3(1), pp. 23-34 (2011)
- Colhon, M. *XML Semantic Schema Annotation for Dependency Relationships*, Annals of the University of Craiova, Mathematics and Computer Science Series, ISSN: 1223-6934, vol. 37(3), pp. 64-70 (2010)
- Colhon, M., Hristea, F. *Syntactic Relations in the Context of Performing Word Sense Disambiguation*, Journal of Knowledge, Communications and Computing Technologies (JKCCT), ISSN: 2067-0958, II(2), pp. 1-9 (2010)
- Colhon, M. *A Contrastive Study of Syntactic Constituents in English and Romanian Texts*, In Proceedings of the Workshop "Language Resources and Tools with Industrial Applications", Adrian Iftene, Diana-Maria Trandabăț (eds.), pag. 11-20, ISSN 1843-911X (2011)
- Colhon, M. *Parallel Treebank From Word-Aligned Bilingual Corpus. Language Engineering for Phrasal Alignments*, In Proceedings of 15th International Conference on System Theory, Control and Computing, Mihail Voicu (ed.), pag. 172-177, ISSN 2068-0465, (2011)
- Colhon, M., Dănciulescu, D. *Semantic Schemas for Natural Language Generation in Multilingual Systems*, Journal of Knowledge, Communications and Computing Technologies (JKCCT), ISSN: 2067-0958, II(1), pp. 10-17 (2010)
- Colhon, M., Țăndăreanu, N. *A Semantic Schema - based Approach for Natural Language Translation*, WSEAS Journal Transactions on Computers, ISSN: 1109-2750, 9(11), pp. 1307 - 1317 (2010)
- Colhon, M., Țăndăreanu, N. *An Approach for Contextual Translation based on Semantic Schemas*, Latest Trends on Computers, vol. I, ISSN: 1792-4251, ISBN: 978-960-474-201-1, pp. 294-299 (2010)
- Ghindeanu, M., Popîrlan, C. *A Natural Language Processing System using*

Java-Prolog Technology, Proceedings of The 5th International Conference on Artificial Intelligence and Digital Communications - AIDC 2005, vol. 105, pp. 63-68, ISBN 973-671-055-6 (2005)

Reprezentări de cunoștințe cu frameuri

Ghindeanu, M., Popîrlan, C. *A Natural Language Processing System using Java-Prolog Technology*, Proceedings of The 5th International Conference on Artificial Intelligence and Digital Communications - AIDC 2005, vol. 105, pp. 63-68, ISBN 973-671-055-6 (2005)

Reprezentări de cunoștințe condiționale:

Țândăreanu, N., Colhon, M., Zamfir, C. *Embedding Conditional Knowledge Bases into Question Answering Systems and Java Implementation*, International Journal of Computers, 4(4), ISSN: 1998-4308, pp. 169-176 (2010)

Țândăreanu, N., Colhon, M., Zamfir, C. *A Spoken Question Answering System Based on Conditional Knowledge* Latest Trends on Computers, vol. I, ISSN: 1792-4251, pp. 220-225 (2010)

Țândăreanu, N., Colhon, M., Zamfir, C. *XML Schema Language Specifications for Conditional Knowledge*, Advances in Communications, Computers, Systems, Circuits and Devices, European Conference of Computer Science (ECCS '10), pp. 149-154, ISSN: 1792-6637 (2010)

Colhon, M., Țândăreanu, N. *The Inference Mechanism in Conditional Schemas*, Annals of the University of Craiova, Mathematics and Computer Science Series, 37(1), ISSN: 1223-6934, pp. 55-70 (2010)

Țândăreanu, N., Ghindeanu, M. *Conditional graphs generated by conditional schemas*, Annals of University of Craiova, Mathematics and Computer Science Series, ISSN: 1223-6934, 36(1), pp. 1-11 (2009)

Țândăreanu, N., Ghindeanu, M. *Towards a Mathematical Modelling of Conditional Knowledge*, Research Notes in Artificial Intelligence and Digital Communications, vol.103, ISBN 973-8419-71-9, Craiova, p.5-15 (2003)

Reprezentarea cunoștințelor incerte

Iancu, I., Constantinescu, N., Colhon M., *Fingerprints Identification using a Fuzzy Logic System*, International Journal of Computers Communications & Control, ISSN 1841- 9836, vol.5(4), pp. 525-531 (2010)

Iancu, I., Colhon, M., Dupac M., *A Fuzzy Controller with Various T-norms Applied in Robot Navigation*, WSEAS Transactions on Systems and Control, ISSN: 1991-8763, 5(11), pp. 882-891 (2010)

Iancu, I., Colhon, M., Dupac, M., *A Takagi-Sugeno Type Controller for Mobile Robot Navigation*, Proceedings of WSEAS International Conf. on Computational Intelligence, pp. 29-34 (2010)

Iancu, I., Colhon, M., *Mamdani FLC with various implications*, Proceedings of 11th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing - SYNASC 2009, published by IEEE Computer Society's Conference Publishing Service (CSP)

Ghindeanu, M., *Fuzzy Controller for Spatial Objects Recognition*, Proceedings of the 6th International Conference on Artificial Intelligence and Digital Communications - AIDC 2006, vol. 106, pp. 118-125 (2006)

Grafuri stratificate etichetate

Dănciulescu, D., Colhon, M. *Splitting the structured paths in stratified graphs. Application in Natural Language Generation*, An. St. Univ. Ovidius Constanța, 22(2) pp. 59-69 (2014) Ghindeanu, M. *Fuzzy Controller for Spatial Objects Recognition*, Research Notes in Artificial Intelligence and Digital Communications, vol. 106, ISBN 973-742-413-1 978-973-742-413-6, pp.118-125 (2006)

Țândăreanu, N., Ghindeanu, M. *Hierarchical Reasoning Based on Stratified Graphs. Application in Image Synthesis*, Proceedings of 15th International Workshop on Database and Expert Systems Applications, Proceedings of DEXA2004, IEEE Computer Society, Zaragoza, Spain, Los Alamitos California, p.498-502 (2004)

Țândăreanu, N., Ghindeanu, M. *Image Synthesis from Natural Language Description*, Research Notes in Artificial Intelligence and Digital Communications, vol.103, ISBN 973-8419-71-9, Craiova, p.82-96 (2003)

Scheme semantic

Colhon, M. *XML Semantic Schema Annotation for Dependency Relationships*, Annals of the University of Craiova, Mathematics and Computer Science Series, ISSN: 1223-6934, vol. 37(3), pp. 64-70 (2010)

Colhon, M., Dănciulescu, D. *Semantic Schemas for Natural Language Generation in Multilingual Systems*, Journal of Knowledge, Communications and Computing Technologies (JKCCT), ISSN: 2067-0958, II(1), pp. 10-17 (2010)

Colhon, M., Țândăreanu, N. *A Semantic Schema - based Approach for Natural Language Translation*, WSEAS Journal Transactions on Computers, ISSN: 1109-2750, 9(11), pp. 1307 - 1317 (2010)

Colhon, M., Țândăreanu, N. *An Approach for Contextual Translation based on Semantic Schemas*, Latest Trends on Computers, vol. I, ISSN: 1792-4251, ISBN: 978-960-474-201-1, pp. 294-299 (2010)

Colhon, M. *Knowledge Transfer and Path-based Reasoning in Hyper-schemas*, Journal of Knowledge, Communications and Computing Technologies, ISSN: 2067-0958, I(1)/2009, pp. 47-58 (2009)

Colhon, M. *Hierarchical Distributed Systems Based on Semantic Schemas*,

- Annals of University of Craiova, Mathematics and Computer Science Series, ISSN: 1223-6934, 36(2), pp. 47-52 (2009)
- Țândăreanu, N., Ghindeanu, M., Nicolescu, S.-A. *Hierarchical Distributed Reasoning System for Geometric Image Generation* International Journal of Computers Communicationa & Control, ISSN 1841- 9836, 4(2), pp. 167-177 (2009)
- Țândăreanu, N., Ghindeanu, M., *Path-based Reasoning in Semantic Schemas*, Annals of University Of Craiova, Mathematics and Computer Science Series, 2008
- Ghindeanu, M., *A Spatial Reasoning HDR System*, Proceedings of the 8th International Conference on Artificial Intelligence and Digital Communications - AIDC 2008, vol. 108, pp. 88-102, ISBN 978-973-671-161-9 (2008)
- M. Ghindeanu, *Constructing Architectures for an Hierarchical Distributed Reasoning System Based on its Inputs*, Proceedings of the 5th International Conference on Cybernetics and Information Technologies, Systems and Applications (CITSA 2008), Orlando, Florida, USA , 2008
- Țândăreanu, N., Ghindeanu, M. *Hierarchical Distributed Reasoning System for Geometric Image Generation*, International Journal of Computers, Communication and Control, 2008
- Țândăreanu, N., Ghindeanu, M. *Hierarchical Semantic Structures Applied in Automatic Image Generation*, Proceedings of the 11th IASTED International Conference on Intelligent Systems and Control (ISC 2008), 2008
- Ghindeanu, M. *The Knowledge Domain of an Hierarchical Distributed System Determines its Architecture*, Annals of University Of Craiova, Mathematics and Computer Science Series, vol. 34, pag. 107-114, ISSN: 1223-6934, 2007
- Ghindeanu, M. *Joining Semantic Schemas in Vision of a Distributed System Reasoning*, Research Notes in Artificial Intelligence and Digital Communications, vol. 107, ISBN 978-973-671-135-0, Craiova, pag. 123-132, 2007
- Țândăreanu, N., Ghindeanu, M. *Properties of Derivations in a Semantic Schema*, Annals of University of Craiova, Mathematics and Computer Science Series, ISSN 1223-6934, vol.33, pp. 147-155 (2006)
- Țândăreanu, N., Ghindeanu, M. *A Three-Level Distributed Knowledge System Based on Semantic Schemas*, Proceedings of 16th International Workshop on Database and Expert Systems Applications, Proceedings of DEXA'05 (TAKMA2005), IEEE Computer Society, Copenhagen, p.423-427, 2005.