



TECHNICAL UNIVERSITY
OF CLUJ-NAPOCA, ROMANIA

Fundamental field: Engineering
Specialisation: Computer Science

HABILITATION THESIS

- ABSTRACT -

Shallow and Deep Understanding of Text Data

Assoc. Prof. Eng. Camelia LEMNARU, Ph.D.
Faculty of Automation and Computer Science
Technical University of Cluj-Napoca

- Cluj-Napoca -
2024



TECHNICAL UNIVERSITY
OF CLUJ-NAPOCA, ROMANIA

This thesis outlines my main academic trajectory since the Ph.D. defence in 2012, focusing on the main research achievements, projecting an accurate professional snapshot of the last 12 years and setting up future research and professional goals. My educational journey at the Technical University of Cluj-Napoca has begun with a bachelor's degree in Computer Science (class of 2006) and continued with a master's degree in Software Engineering (2006-2008), culminating with a Ph.D. in 2012. Since 2007, I have been a member of the Computer Science Department and the Knowledge Engineering Group (KEG). During the last 17 years, I've contributed to the development of our department, by engaging in various teaching, research, and departmental activities. Beyond academia, I've been actively involved in the Romanian Association on AI and in drafting a national strategic framework for AI.

Towards the end of my Ph.D. period, I have focused on tackling data-imbalance with cost-sensitive learning techniques. For a short period after my Ph.D. defence, I have continued to work on methods for characterizing and alleviating the effects of class imbalance, but in the following years I have slowly shifted focus to Natural Language Processing, a domain which has produced the most relevant results of my recent research efforts and represents the subject of the current thesis. The main objective during this period was to work towards enhancing machine understanding of text, covering syntactic structures, text classification, and delving into deeper comprehension with AMR parsing. I've tackled challenges such as multilingualism, multitask learning, class imbalance, and interpretability across various NLP areas. The great majority of my *recent* publications are in this area:

- 5 journal publications (all ISI, 3 Q1/Q2)
- 44 conference proceeding publications, out of which 36 ISI proceedings, 1 in a workshop at ECML (A* conference), 1 at EACL (A* conference), 1 at ICDE (A* conference)
- 4 book chapters

Additionally, throughout my entire academic career, I have been involved in four research grants as lead and six as team member, I have acted as reviewer and program committee member for several international conferences (ECML-PKDD, EMNLP, ICCP, CoDiT, etc) and I act as a regular reviewer of several international ISI journals (IEEE JHBI, IEEE/ACM TASL, IJNS, Sensors, etc). I have been involved in the organization of several scientific events and I have been a member of mentoring/supervision committees of several Computer Science Ph.D. students.

In teaching, I have been involved in planning, preparing and delivering lecture and practical activities (seminar and laboratory sessions) for several fundamental subjects: Data Structures and Algorithms (1st year, bachelor), Fundamental Algorithms (2nd year, bachelor), Logic Programming (3rd year, bachelor) and Big Data in Security (1st year, master). Tightly coupled with my teaching efforts, I am actively involved in bachelor and master theses elaboration: I coordinate, on average, between 7 and 9 bachelor theses per year, either jointly with my KEG colleagues, or individually; additionally, I advise on average 3-4 master students per year. The main topics are related to NLP research, but – especially for the bachelor theses – I adapt the topic to suit the interest and expertise level of the student. For all the subjects I teach, I have developed new teaching resources which are constantly updated and improved.



TECHNICAL UNIVERSITY
OF CLUJ-NAPOCA, ROMANIA

In addition to my research and teaching activities, I am involved in various activities undergone at department and university level: ACM International Collegiate Programming Contest (ACM-ICPC) teams coordinator, Erasmus Computer Science Department coordinator, schedule of classes responsible, student internship tutor, assistant year advisor (4th year, Information Technology).

On the medium term, my academic focus is on diversifying the department's educational offer by developing a new NLP master course, increasing student engagement and performance as well as their interest in undergoing a mobility for studies (in general, and in one of the EuT+ partners in particular). The research objectives for the following period focus on strengthening the NLP capabilities of the KEG research group at UTCN, by attracting at least 1 new permanent member and several Ph.D. students, increasing the visibility and impact of the group, via publication (and participation) at top-tier journals and conferences and contributing to increasing the number of research grants (in partnership with both academia and industry).

Thank you.