UNIVERSITAS CAROLINA

FACULTAS MATHEMATICAE PHYSICAEQUE DISCIPLINAE

STUDY GUIDE
Bachelor's and Master's Programmes
2020/2021

Master of Computer Science

Study started in 2020 and later

General Information

Study programs and their specializations

1. Computer Science - Discrete Models and Algorithms
 – Discrete mathematics and algorithms
 – Geometry and mathematical structures of computer science
 – Optimization
2. Computer Science - Theoretical Computer Science
3. Computer Science - Software and Data Engineering
 – Software engineering
 – Software development
 – Web engineering
 – Database systems
 – Big data processing
4. Computer Science - Software Systems
 – System programming
 – Dependable systems
 – High performance computing
5. Computer Science – Language Technologies and Computational Linguistics
 – Computational and formal linguistics
 – Statistical and machine learning methods in Natural Language Processing
6. Computer Science - Artificial Intelligence
 – Intelligent agents
 – Machine learning
 – Robotics
7. Computer Science – Visual Computing and Game Development
 – Visual Computing
 – Computer game development

While your study program has been specified already in your application, the definitive choice of your specialization (where applicable) is made only later, when enrolling for the state final exam.

Computer science is a dynamically evolving discipline, and therefore we are constantly adapting the content of our study programs to important new trends. In their own interest, students should keep track of the current state of study plans as the list

of offered courses may be expanded and modified, or other minor changes may take place.

Assumed knowledge

It is assumed that an incoming student has a sufficient knowledge of mathematics, theoretical computer science, and programming. In particular, students are expected to have a good knowledge of mathematics at the level of our bachelor courses NMAI054 Mathematical Analysis 1, NMAI058 Linear Algebra 2, NMAI059 Probability and Statistics 1.

Students are also expected to have knowledge equivalent to the courses NDMI002 Discrete Mathematics, NTIN060 Algorithms and Data Structures 1, NTIN061 Algorithms and Data Structures 2, NTIN071 Automata and Grammars, and NAIL062 Propositional and Predicate Logic. Knowledge from these courses is also expected for the state final exam.

We also expect students to have good knowledge of programming at least at the level of our bachelor courses NPRG030 Programming 1 and NPRG031 Programming 2.

Students who are missing knowledge in some of the above-mentioned areas are advised to consider taking the relevant bachelor courses in the first year of their Master's studies. Please do not hesitate to contact the program coordinator in case of doubt.

If a student has successfully completed one of the obligatory or optional courses of their study program during his/her previous bachelor's study at the Faculty of Mathematics and Physics, they may apply for recognition of the fulfillment of these obligations. A student coming to the Faculty after obtaining a bachelor's degree at another university may apply for recognition of obligatory or optional courses on the basis of previous completion of a similar subject. The awarding of credits for courses completed in the bachelor's study is regulated by Article 12 of the Rules for the Organization of Studies at the Faculty of Mathematics and Physics.

Team project

Study plans of master programs in the study area Computer Science offer the possibility of participation in a team project. In the study programs Software Systems, Software and Data Engineering, Visual Computing and Game Development the team project is obligatory, while in programs Artificial Intelligence, Language Technologies and Computational Linguistics the team project is elective. The student chooses one out of three types of a team project: a Software project, a Research project, or a Company Project. The software project is a classic student project, where a team of 3-6 students develops a larger software system. The research project allows a student to temporarily join an existing research team at the faculty, in which the student works on a particular research/development task. The company project allows a student to accomplish the team software project outside of the faculty environment, in a company, while still meeting the standards set commonly for all project types. In case of a more difficult project assignment, extra credits can be awarded using the course Increased project scope. Approval and evaluation of projects is guided by the regulations of the respective study program coordinator.

State Final Exam

The state final exam consists of two parts: a defense of the Master's (diploma) thesis, and an oral examination. The student can enroll for each part separately. To finish the studies, both parts of the state final exam must be completed successfully.

Requirements to enroll for the state final exam

- obtaining at least 120 credits,
- passing all obligatory courses of a given study program,
- obtaining a given number of credits from the elective courses of a given program and specialization,
- submitting the Master's thesis by the specified deadline (for defense of the Master's thesis).

Master's (diploma) thesis

Students are advised to select the topic of their Master's (diploma) thesis at the end of the first semester. The faculty departments offer many topics for Master's theses each year, and students may also suggest their own topics. We recommend selecting the topic of your thesis primarily from those offered by the department that coordinates your study program. If you prefer a topic offered by another department or your own topic, please consult the coordinator of your study program.

After the topic is assigned, the student enrolls in the following obligatory courses:

Code	Subject	Credits	Winter	Summer
NSZZ023	**Diploma Thesis I**	6	—	0/4 C
NSZZ024	**Diploma Thesis II**	9	0/6 C	—
NSZZ025	**Diploma Thesis III**	15	—	0/10 C

Course credits for these courses are granted by the thesis advisor on the basis of student's work on the thesis. We suggest the students to discuss with their advisor the expected amount of work and the milestones for each of these courses. All three courses can be enrolled in both winter and summer semesters.

Oral examination

The oral part of the state final exam has a similar structure for all study programs. The student is examined from several obligatory and several optional examination areas specific to a given study program and selected specialization. The student will select these optional examination areas when registering for the final exam. A more detailed description can be found in the relevant section of each study program.

Note that not all the courses are available in English every year. We recommend students to contact the study program coordinator for the selected study program and discuss individual study plans prior the beginning of each semester.

5 Computer Science – Language Technologies and Computational Linguistics

Coordinated by: Institute of Formal and Applied Linguistics
Study programme coordinator: Doc. RNDr. Markéta Lopatková, Ph.D.

Specializations:

- Computational and formal linguistics
- Statistical and machine learning methods in Natural Language Processing

The graduate is familiar with mathematical and algorithmic foundations of automatic natural language processing, with theoretical foundations of formal description of natural languages, as well as with state-of-the-art machine learning techniques. The student acquires the skills in designing and development of systems to automatically process large quantities of language data, written and spoken, structured and unstructured alike, and to solve language-related tasks, such as information retrieval, question answering, summarization and information extraction, machine translation, and speech processing.

The graduate is well prepared for doctoral studies in computational linguistics and language technologies, as well as for a professional career in the public or private sector. Given the general applicability of machine learning and data driven methods, the graduate is well equipped to use these methods not only in natural language processing tasks but also in other domains where large quantities of both structured and unstructured data are being analyzed (finances, economy, biology, medicine, and other domains). The student acquires programming experience and soft skills required for team work on applications that involve machine learning or human-computer interaction.

5.1 Obligatory courses

Code	Subject	Credits	Winter	Summer
NTIN066	**Data Structures 1**	6	—	2/2 C+Ex
NTIN090	**Introduction to Complexity and Computability**	4	2/1 C+Ex	—
NPFL063	**Introduction to General Linguistics**	4	2/1 C+Ex	—
NPFL067	**Statistical Methods in Natural Language Processing I**	5	2/2 C+Ex	—
NPFL114	**Deep Learning**	7	—	3/2 C+Ex
NSZZ023	**Diploma Thesis I**	6	—	0/4 C
NSZZ024	**Diploma Thesis II**	9	0/6 C	—
NSZZ025	**Diploma Thesis III**	15	—	0/10 C

5.2 Elective courses - Set 1

The student needs to obtain at least 40 credits in total for the elective courses. Of these 40 required credits, at most 6 credits can be obtained from project courses (set 2 below) and at most 10 credits from the additional set of elective courses (set 3 below).

Code	Subject	Credits	Winter	Summer
NPFL006	Introduction to Formal Linguistics	3	2/0 Ex	—
NPFL038	Fundamentals of Speech Recognition and Generation	5	2/2 C+Ex	—
NPFL068	Statistical Methods in Natural Language Processing II	5	—	2/2 C+Ex
NPFL070	Language Data Resources	4	1/2 MC	—
NPFL075	Dependency Grammars and Treebanks	5	—	2/2 C+Ex
NPFL079	Algorithms in Speech Recognition	5	—	2/2 C+Ex

NPFL082	Information Structure of Sentences and Discourse Structure	2	—	0/2 C
NPFL083	Linguistic Theories and Grammar Formalisms	5	—	2/2 C+Ex
NPFL087	Statistical Machine Translation	5	—	2/2 C+Ex
NPFL093	NLP Applications	4	—	2/1 MC
NPFL094	Morphological and Syntactic Analysis	3	2/0 MC	—
NPFL095	Modern Methods in Computational Linguistics	3	0/2 C	—
NPFL097	Unsupervised Machine Learning in NLP	3	1/1 C	—
NPFL099	Statistical Dialogue Systems	4	2/1 C+Ex	—
NPFL100	Variability of Languages in Time and Space	2	1/1 C	—
NPFL103	Information Retrieval	5	2/2 C+Ex	—
NPFL104	Machine Learning Methods	4	—	1/2 C+Ex
NPFL122	Deep Reinforcement Learning	5	2/2 C+Ex	—
NPFL128	Language Technologies in Practice	4	—	2/1 MC

5.3 Elective courses - Set 2 (project courses)

The student can select at most one of the project courses as an elective course; at most 6 credits count as credits for elective courses. (Other potential credits for courses from this set count as credits for free courses.)

Code	Subject	Credits	Winter	Summer
NPRG069	Software Project	12	0/8 C	0/8 C
NPRG070	Research Project	9	0/6 C	0/6 C
NPRG071	Company Project	6	0/4 C	0/4 C

5.4 Elective courses - Set 3 (additional courses)

The student can select any course from the following set of additional courses; at most 10 credits count as credits for elective courses. (Other potential credits for courses from this set count as credits for free courses.)

Code	Subject	Credits	Winter	Summer
NAIL025	Evolutionary Algorithms 1	5	2/2 C+Ex	—
NAIL069	Artificial Intelligence 1	4	2/1 C+Ex	—
NAIL070	Artificial Intelligence 2	3	—	2/0 Ex
NAIL104	Probabilistic graphical models	3	2/0 Ex	—
NPGR036	Computer Vision	5	—	2/2 C+Ex

5.5 State Final Exam

The state final exam for the program Language Technologies and Computational Linguistics consists of one obligatory examination area for both specializations (examination area 1), one obligatory area dependent on the selected specialization (examination area 2 or examination area 3), and one elective examination area (examination

areas 4 and 5). As the last examination area, the student may also select the obligatory area of the other specialization of this study program. In total, each student gets questions from three examination areas.

Examination areas

1. Fundamentals of natural language processing (obligatory for both specializations)
2. Linguistic theories and formalisms (obligatory for the specialization Computational and formal linguistics)
3. Statistical methods and machine learning in computational linguistics (obligatory for the specialization Statistical and machine learning methods in Natural Language Processing)
4. Speech, dialogue and multimodal systems (elective)
5. Applications in natural language processing (elective)

Knowledge requirements

1. Fundamentals of natural language processing

Phonetics, phonology, morphology, syntax, semantics, pragmatics. Ambiguity, arbitrariness. Description and prescription. Diachronic and synchronic language description. Fundamentals of information theory. Markov models. Language modeling and smoothing. Word classes. Annotated corpora. Design and evaluation of linguistic experiments, evaluation metrics. Morphological disambiguation and syntactic analysis. Basic classification and regression algorithms.

Recommended courses

Code	Subject	Credits	Winter	Summer
NPFL063	Introduction to General Linguistics	4	2/1 C+Ex	—
NPFL067	Statistical Methods in Natural Language Processing I	5	2/2 C+Ex	—

2. Linguistic theories and formalisms

Functional Generative Description. Prague Dependency Treebank. Universal Dependencies. Other grammar formalisms (overview and basic characteristics). Phonetics, phonology. Computational Morphology. Surface and deep syntactic structure; valency. Computational lexicography. Topic-focus articulation; information structure, discourse. Coreference. Linguistic typology. Formal grammars and their application in rule-based morphology. Parsing.

Recommended courses

Code	Subject	Credits	Winter	Summer
NPFL063	Introduction to General Linguistics	4	2/1 C+Ex	—
NPFL006	Introduction to Formal Linguistics	3	2/0 Ex	—
NPFL075	Dependency Grammars and Treebanks	5	—	2/2 C+Ex
NPFL083	Linguistic Theories and Grammar Formalisms	5	—	2/2 C+Ex
NPFL094	Morphological and Syntactic Analysis	3	2/0 MC	—

3. Statistical methods and machine learning in computational linguistics

Generative and discriminative models. Supervised machine learning methods for classification and regression (linear models, other methods: naive Bayes, decision trees, instance-based learning, SVM and kernels, logistic regression). Unsupervised machine learning methods. Language models, noisy channel model. Model smoothing, model combination. HMM, trellis, Viterbi, Baum-Welch. Algorithms for statistical tagging. Algorithms for constituency and dependency statistical parsing. Neural networks in machine learning. Convolution and recurrent networks. Word embeddings.

Recommended courses

Code	Subject	Credits	Winter	Summer
NPFL067	Statistical Methods in Natural Language Processing I	5	2/2 C+Ex	—
NPFL114	Deep Learning	7	—	3/2 C+Ex
NPFL068	Statistical Methods in Natural Language Processing II	5	—	2/2 C+Ex

4. Speech, dialogue and multimodal systems

Fundamentals of speech production and perception. Methods of speech signal processing. HMM acoustic modeling of phonemes. The implementation of the Baum-Welch and Viterbi algorithms in speech recognition systems. Neural models for speech. Methods of speech synthesis. Speech applications. Basic components of a dialogue system. Natural language understanding in dialogue systems. Dialogue state tracking. Methods for dialogue management. User simulation. End-to-end neural dialogue systems. Open-domain dialogue system architectures. Natural language generation. Dialogue systems evaluation. Visual dialogue and multimodal systems.

Recommended courses

Code	Subject	Credits	Winter	Summer
NPFL038	Fundamentals of Speech Recognition and Generation	5	2/2 C+Ex	—
NPFL079	Algorithms in Speech Recognition	5	—	2/2 C+Ex
NPFL099	Statistical Dialogue Systems	4	2/1 C+Ex	—

5. Applications in natural language processing

Spell-checking and grammar-checking. Machine translation. Machine-aided translation. Statistical methods in machine translation. Quality evaluation of machine translation. Speech translation. Information retrieval, models for information retrieval. Query expansion and relevance feedback. Document clustering. Duplicate detection and plagiarism detection. Information retrieval evaluation. Sentiment analysis. Toolkits (GATE, NLTK, NLPTools, Lucene, Terrier).

Recommended courses

Code	Subject	Credits	Winter	Summer
NPFL087	Statistical Machine Translation	5	—	2/2 C+Ex
NPFL093	NLP Applications	4	—	2/1 MC
NPFL103	Information Retrieval	5	2/2 C+Ex	—

				Artificial Intelligence
NPFL128	Language Technologies in Practice	4	—	2/1 MC