

МИНОБРНАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«Новосибирский национальный исследовательский государственный университет»
(Новосибирский государственный университет, НГУ)
**Структурное подразделение Новосибирского государственного университета –
Специализированный учебно-научный центр Университета (СУНЦ НГУ)**
Министерство науки и высшего образования Российской Федерации

СОГЛАСОВАНО Заместитель директора по УР  (Петровская О.В.) 23 ноября 2023 г.	УТВЕРЖДЕНО На заседании ученого совета СУНЦ НГУ Протокол № 48 от 23 ноября 2023 г.	УТВЕРЖДАЮ Директор СУНЦ НГУ  (Некрасова Л.А.) 23 ноября 2023 г.
--	--	--

РАБОЧАЯ ПРОГРАММА

**курса внеурочной деятельности «Введение в математическую и компьютерную
лингвистику»**

Заведующий кафедрой математических наук
Миронов Андрей Евгеньевич, д.ф.-м.н., чл.корр РАН



Новосибирск 2023

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Математическая и компьютерная лингвистика является крайне востребованной и активно развивающейся областью исследований. Это обусловлено задачами искусственного интеллекта и анализа данных, связанными с обработкой информации на естественных языках (как письменной, так и устной). Обработка естественного языка (natural language processing, NLP) в настоящее время опирается преимущественно на методы машинного обучения и статистического анализа больших объемов текстовой информации. В то же время, работающие в этой сфере ИТ-индустрии специалисты отмечают назревшую необходимость более глубокого понимания строения и принципов функционирования естественных языков для решения наиболее сложных задач NLP. В первую очередь, речь идет о семантике предложения. Содержание курса как раз составляет описание моделей такого типа, сочетающих математическую глубину и формальную строгость с возможностью практической реализации.

Важную роль в определении структуры и содержании курса играет сочетание различных моделей языков (в первую очередь, естественных): математических моделей, основанных на формальной семантике Монтегю и синтаксической категориальной грамматике Ламбека, и компьютерных моделей, использующихся при решении конкретных задач NLP и опирающихся на довольно далекие от указанных выше модели и методы. Основными методами здесь являются дистрибутивная семантика (Word2Vec и т.п.) и машинное обучение. Данная область исследований развивается крайне активно. В курс включено описание подхода, предложенного группой исследователей, возглавляемой проф. Б.Коке из Университета Оксфорда. Этот подход интересен тем, что сочетает принцип композициональности Фреге с методами дистрибутивной семантики.

Особенностью школьного курса является раздел, посвященный олимпиадным задачам по лингвистике. Многие из этих задач неявно опираются на рассматриваемые в курсе математические модели и методы анализа текстовой информации.

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ

Слушатели курса должны ознакомиться с с типами и спецификой олимпиадных задач по лингвистике, важнейшими формальными языками и исчислениями с точной семантикой, классическими математическими моделями для естественных языков, современными моделями и методами компьютерной лингвистики для задач обработки текстовой информации.

СОДЕРЖАНИЕ СПЕЦКУРСА

Курс состоит из четырех разделов, посвященных, соответственно, олимпиадным задачам по лингвистике, формальным языкам и логическим исчислениям, математическим моделям для естественных языков, и моделям и методам компьютерной лингвистики.

ТЕМАТИЧЕСКОЕ ПЛАНИРОВАНИЕ

№ п/п	Наименование разделов и тем программы	Количество часов	Воспитательный компонент
Раздел 1. Олимпиадные задачи по лингвистике			
1.1.	Обзор основных типов лингвистических олимпиадных задач и общих методов их решения	4	Развитие и поддержка одаренности обучающихся и обеспечение участия в олимпиадах и конкурсах.
1.2.	Разбор избранных олимпиадных задач	6	
Итого по разделу		10	
Раздел 2. Формальные языки и логические исчисления			
2.1.	Логика высказываний, логика предикатов и темпоральная логика: синтаксис и семантика	4	Готовность осуществлять проектную и исследовательскую деятельность индивидуально и в группе.
2.2.	Иерархия Хомского классов порождающих грамматик и формальных языков	4	
2.3.	Иерархия конечных автоматов	2	
Итого по разделу		10	
Раздел 3. Математические модели для естественных языков			
3.1.	Принцип композициональности Фреге, категориальные грамматики и функциональная семантика Монтегю	4	Установление доверительных отношений между руководителем объединений и обучающимися и между обучающимися непосредственно через беседы, дискуссии.
3.2.	Темпоральная система Райхенбаха и интервальная семантика глагольных времен	4	

3.3.	Обобщенные кванторы и конструктивные модели	2	
Итого по разделу		10	
Раздел 4. Модели и методы компьютерной лингвистики			
4.1.	Дистрибутивная семантика для задач NLP	6	Развитие и поддержка одаренности обучающихся и обеспечение участия в олимпиадах и конкурсах.
4.2.	Модель DisCoCat и ее реализации	4	
Итого по разделу		10	
Итого		40	

Приложение 1.

УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

Список литературы

1. Barbara H. Partee, Alice ter Meulen, and Robert E. Wall, *Mathematical Methods in Linguistics*, Kluwer Academic Publishers, 1990.
2. *Compositionality in Formal Semantics, Selected Papers by Barbara H. Partee*, Blackwell Publishing, 2004.
3. Dawid R. Dowty, Robert E. Wall, and Stanley Peters, *Introduction to Montague Semantics*, D. Reidel Publishing Company, 1989.
4. Horst Lohnstein, *Formale Semantik und Natürliche Sprache*, Springer Fachmedien Wiesbaden GmbH, 1996.
5. Paul H. Portner, *What is Meaning?*, Blackwell Publishing, 2005.
6. И.А.Герасимова, *Формальная грамматика и интенциональная логика*, М., ИФ РАН, 2000.
7. Harry R. Lewis, Christos H. Papadimitriou, *Elements of the Theory of Computation*, Prentice-Hall, 1998.