

**Конференция к 20-летию НКРЯ.  
Круглый стол «Корпусные платформы для корпусов разного типа»**

**Организаторы**

Сергей Александрович Гладилин, с.н.с. ИППИ РАН

Анастасия Дмитриевна Козеренко, с.н.с. Отдела экспериментальной лексикологии и лексикографии ИРЯ РАН

**Тематика**

В рамках круглого стола предполагается обсудить подготовку, разметку и индексирование корпусов различного размера и структурной сложности: от малых корпусов с глубокой ручной разметкой до гигантских мега-корпусов, формируемых на базе сети Интернет. Корпусные платформы НКРЯ и ГИКРЯ будут обсуждаться как примеры, ориентированные на существенно различные виды корпусов. Планируется обсудить различные аспекты – от особенностей применяемого программного обеспечения и организации поисковых систем до взаимодействия между собой различных групп исследователей, формирующих команду корпуса. Круглый стол будет интересен специалистам, участвующим в создании доступных для пользователей в сети Интернет языковых корпусов.

**Обоснование важности**

В настоящее время корпуса, доступные для пользователей через сеть Интернет, являются одним из важнейших инструментов корпусной лингвистики. Их создатель берет на себя задачи по подготовке и хранению корпусных данных, обеспечению их инструментами поиска и лингвистического анализа, обеспечения удобного интерфейса для доступа к этим инструментам, поддержания и своевременного обновления программного и аппаратного обеспечения, необходимого для функционирования системы. Все это существенно снижает порог вхождения для исследователей, позволяя им приступить к исследованиям без разворачивания корпуса и инструментов к нему на собственном персональном компьютере. В то же время такие корпусные системы с открытым доступом через Интернет сами по себе приобретают высокую сложность: они включают в себя многочисленные и/или крупные корпуса текстов, одновременный доступ к которым многочисленных пользователей из сети должен быть обеспечен аппаратными ресурсами и осуществляться через специальное оптимизированное программное обеспечение. В настоящее время двумя наиболее значимыми корпусными проектами для русского языка в России являются НКРЯ и ГИКРЯ. Это делает актуальным проведение круглого стола для обмена опытом между этими двумя корпусными платформами с привлечением специалистов из других проектов.