**Проектное предложение**

|  |  |
| --- | --- |
| Тип проекта | Исследовательский |
| Название проекта | *Имя исследователя: устранение неоднозначности библиометрических данных* |
| Подразделение инициатор проекта | Отдел исследований человеческого капитала |
| Руководитель проекта | Качанов Юрий Львович |
| Описание содержания проектной работы | Адекватное устранение неоднозначности имен авторов в библиометрических базах данных является предварительным условием для проведения наукометрического анализа. Это нетривиальная задача, поскольку разные авторы могут иметь одно и то же имя (омонимы), а один автор может публиковаться под разными именами (синонимами). В случае сколько-нибудь масштабных библиометрических исследований невозможно устранить неоднозначность идентификации каждого автора вручную. В литературе было предложено несколько подходов для устранения неоднозначности имени автора, но их корректная статистическая оценка еще не была реализована. В настоящем проекте планируется |
| Цель и задачи проекта | *Цель***:** Разработка метода устранения неоднозначности имени автора публикации без учителя.  *Задачи*:   * конструирование модели для оценки сходства априорных упоминаний автора вместо обучения модели с размеченными данными; * обзор мирового опыта идентификации автора научной публикации; * сравнительный анализ различных методов идентификации автора научной публикации. |
| Виды деятельности, выполняемые студентом в проекте/отрабатываемые навыки | * Анализ существующих в литературе подходов к сравнению методов идентификации автора научной публикации. * Программирование алгоритма идентификации автора на языке Python. * Анализ библиометрических данных базы INSPIRE. * Подготовка частей аналитического отчёта на основе литературного обзора и результатов статистической обработки данных. |
| Сроки реализации проекта | с 26 марта 2021 по 28 июня 2021 |
| Количество кредитов | 4 кредита |
| Тип занятости студента | Удалённая работа с регулярными отчётами |
| Интенсивность (часы в неделю) | 8 часов в неделю |
| Вид проектной деятельности | Индивидуальная |
| Требования к студентам, участникам проекта | 1. Интерес к количественным методам социальных наук. 2. Интерес к математическому моделированию в социальных науках. 3. Базовые знания теории вероятностей и математической статистики. 4. Навыки программирования на Python. 5. Готовность к научному поиску, деятельности в условиях неопределённости. 6. Приветствуется наличие опыта работы с библиометрическими базами данных. |
| Планируемые результаты проекта | Аналитический отчёт с чётко выделенными индивидуальными частями каждого из участников проекта |
| Формат представления результатов, который подлежит оцениванию (отчет студента по проекту) | Часть текста аналитического отчета |
| Критерии оценивания результатов проекта | * Степень вовлеченности в работу по проекту * Творческий подход к исследованию * Поиск литературы * Корректность интерпретации источников по теме исследования * Результаты работы с данными * Уровень подготовки аналитического отчёта |
| Количество вакантных мест на проекте | 5 |
| Критерии отбора студентов в проект (применяются в случае большого количества заявок на проект) | * Интерес к теме проекта * Предрасположенность к научно-исследовательской деятельности * Успеваемость по профессиональным дисциплинам |
| Образовательные программы | Студенты ОП «Социология», «Бизнес-информатика», «Информатика и вычислительная техника», «Компьютерные науки и анализ данных», «Прикладная математика». |
| Территория | Отдел исследований человеческого капитала |