

Секция «Социология социальных сетей и новые методы анализа общественных процессов»

Разработка автоматического универсального инструмента для осуществления систематического подхода к обзору литературы на примере темы эко-практик

Научный руководитель – Ротмистров Алексей Николаевич

Марченко Анастасия Васильевна

Студент (магистр)

Национальный исследовательский университет «Высшая школа экономики», Факультет социальных наук, Москва, Россия
E-mail: avmarchenko_2@edu.hse.ru

Обзор научной литературы является одним из ключевых этапов для проведения любого исследования. Поиск и анализ публикаций по релевантной для исследователя тематике дает возможность понять, насколько та или иная проблема уже изучена, и обнаружить «пробелы» в имеющейся системе знаний. На данный момент можно выделить две стратегии для подготовки литературного обзора: (1) экспертный отбор литературы (качественный анализ публикаций по изучаемой теме) и (2) алгоритмический отбор литературы (библиометрический анализ публикаций с использованием показателей цитирования) [3]. Именно вторая стратегия позволяет избежать потери важных для обзора работ и является более прозрачной с точки зрения обоснования выбора конкретных источников как наиболее значимых для изучения поставленной исследовательской проблемы.

Библиометрический анализ [5] как инструмент для составления научных обзоров все чаще встречается в литературе. Авторы, использующие данный инструмент, стараются разработать новые методики или же включить в процедуру анализа дополнительные этапы, которые бы упростили ее или повысили качество отбора и анализа источников [1, 4, 6]. Однако подготовка литературного обзора все еще считается трудоемкой задачей. Кроме того, нет инструмента, который был бы универсальным и мог бы автоматически выполнить все этапы (от сбора источников до построения сети) систематического обзора литературы по любой тематике.

В качестве предметной области, на примере которой будет изучен и продемонстрирован систематический подход к обзору литературы, будет использована активно развивающаяся в последнее десятилетие тема экологических практик. Данный выбор объясняется следующими условиями. Во-первых, в обзоре глобальных рисков за 2022 год первые три позиции занимают риски, связанные с экологией [8]. Во-вторых, основным фокусом программ, направленных на достижение целей устойчивого развития в области экологии, является формирование эко-практик среди населения [7]. И, в-третьих, общее количество публикаций на тему эко-практик из разных областей науки с каждым годом заметно увеличивается [на основании составленных авторами расчетов за последние 10 лет в базе данных научного цитирования Scopus]. Исходя из этого, можно предположить, что появляются новые связи между авторами, предметами исследований и условиями для их изучения.

Таким образом, в рамках данного исследования возникает два ключевых исследовательских вопроса. Первый (методологический): с использованием каких методов можно автоматизировать и универсализировать систематический подход к обзору литературы? Второй (содержательный): каковы основные тенденции в изучении экологических практик с концептуальной и методологической точек зрения?

Для поиска ответа на поставленные исследовательские вопросы планируется проанализировать источники на тему экологических практик в базе данных научного цитирования

Scopus и в научной интегрированной с РИНЦ библиотеке eLibrary. Систематический подход к обзору литературы подразумевает осуществление трех этапов [1]:

- (1) Сбор публикаций по релевантной для исследователя теме (скачивание базы вручную);
- (2) Анализ собранных публикаций с целью выделения ключевых слов, соответствующих релевантной для исследователя теме, для повторного сбора публикаций и их библиографической информации (с помощью текст-майнинга);
- (3) Построение сети и ее интерпретация (с помощью сетевого анализа).

В рамках данной работы эта трехэтапная методика будет взята за основу для разработки автоматического и универсального инструмента в среде программирования Python. На первом этапе вместо сбора полных текстов публикаций вручную будет осуществлен сбор аннотаций публикаций по релевантной для исследователя теме. Использование в качестве единицы анализа не полного текста публикации, а только ее аннотации объясняется тем, что она дает полное представление о проведенном исследовании и является максимально информативной и точной. Для упрощения и автоматизации процесса сбора будет использована процедура веб-скрапинга - автоматического извлечения данных из веб-ресурсов [2]. Второй и третий этап останутся без существенных изменений за исключением того, что на втором этапе текст-майнинг будет применен именно к аннотациям, а не к полным текстам найденных публикаций.

Таким образом, в рамках данной работы планируется, во-первых, разработать инструмент (в виде скрипта на языке программирования Python), который позволит автоматически осуществлять систематический обзор литературы на любую релевантную для авторов тему. Полученный инструмент позволит упростить и сделать более эффективной процедуру анализа и отбора научной литературы. Он будет полезен исследователям из любой научной области, а также заинтересованным лицам из коммерческих и практико-ориентированных секторов. А, во-вторых, выделить ключевых авторов и ключевые концепции и методы для изучения экологических практик на момент проведения исследования.

Источники и литература

- 1) Гибсон Э., Дайм Т., Гарсес Э., Дабич М. Библиометрический анализ как инструмент выявления распространенных и возникающих методов технологического Форсайта // Форсайт. 2018. Т. 12. № 1. С. 6–24.
- 2) Жучкова С.В., Ротмистров А.Н. Автоматическое извлечение текстовых и числовых веб-данных для целей социальных наук // Социология: методология, методы, математического моделирование. 2020. № 50–51. С. 141–183.
- 3) Мальцева Д.В., Моисеев С.П. Отбор источников для систематического обзора литературы: сравнение экспертного и алгоритмического подходов // Социология: методология, методы, математического моделирование. 2018. № 47. С. 7–43.
- 4) Мальцева Д.В., Щеглова Т.Е., Ващенко В.А., Моисеев С.П. Здоровые города: Выделение актуальных трендов исследований в научной литературе и социальных медиа // Социодиггер. 2022. Т. 3. № 9. С. 41–59.
- 5) Bar-Ilan J. Informetrics at the Beginning of the 21st Century – A Review // Journal of Informetrics. 2008. Vol. 2. No. 1. P. 1–52.
- 6) Batagelj V., Doreian P., Ferligoj A., Kejzar N. Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution. Chichester: Wiley. 2014.
- 7) Sustainable Development Goals // United Nations. 2022. URL: <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>

- 8) The Global Risks Report 2022, 17th Edition // World Economic Forum. 2022. URL: <https://www.weforum.org/reports/global-risks-report-2022>