

ЦИФРОВЫЕ СЛЕДЫ ОБРАЗОВАТЕЛЬНОЙ МИГРАЦИИ

Н.К. Габдрахманов, Т.С. Бабкина
Национальный исследовательский университет
«Высшая школа экономики»,
Институт образования

Аннотация

Использование цифровых следов является мощным драйвером исследований новой формации. Пока использование этих данных сильно ограничено и публикаций на данную тему не так много. В данной работе авторами делается попытка использования данных цифровых следов и геоинформационных систем для анализа и интерпретации образовательной миграции в регионах России.

Введение

Вопросы миграции населения привлекают внимание, как научного сообщества, так и общественности [1]. С процессом миграции связаны многие социально-экономические процессы, в частности движение рабочей силы и связанное с этим неравномерное распределение плотностей в пространстве [2] и т.д. Первым серьезным пиком в динамике миграции населения является возраст 18 лет, когда выпускники школ и их родители стоят перед выбором высшего учебного заведения и дальнейшей карьеры [3]. Это ситуация характерная не только для России, но и для мира в целом (Рис. 1). Где-то этот процесс проходит интенсивнее, где-то имеет меньший масштаб. Второй пик, не меньший по интенсивности наблюдается в возрасте выпуска из университета и определения места приложения труда. И в первом и во втором случае важно разграничивать внутрирегиональную и межрегиональную миграцию. Они имеют различные интенсивности и масштабы.

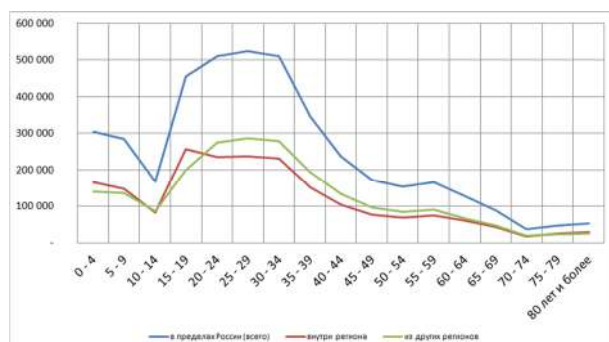


Рис.1. Возрастной состав мигрантов по РФ в 2017 г.

Учет миграции ведется на основе листов прибытия и выбытия, а также на основании данных переписи населения. Данная методология имеет свои достоинства и недостатки. В тоже время, каждый человек, который пользуется современными девайсами и социальными сетями оставляет множество цифровых следов [4], которые являются прекрасным источником для сбора информации о перемещениях населения в пространстве.

Использование цифровых следов в анализе миграционных потоков

Социальная сеть «ВКонтакте» [5], наиболее распространенная в РФ, дает возможность расширить представление об образовательной миграции. Открытые данные пользователей социальной сети содержат информацию о школах, местах получения высшего образования и текущего трудоустройства. Преимущество в использовании данных из социальной сети «ВКонтакте» достигается за счет возможности выявления индивидуальных образовательных траекторий. Функционал социальной сети позволяет про-

водить такое исследование в анонимном виде. Например, на Рис. 2 представлена карта образовательных миграций, связанных с поступлением в один из перечисленных университетов: Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ), Национальный исследовательский Томский государственный университет (ТГУ), Северный (Арктический) федеральный университет (САФУ), Северо-Восточный федеральный университет (СВФУ), Уральский федеральный университет (УрФУ), Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М.В. Ломоносова» (МГУ).



Рис.2. Студенческие миграции: поступление в НИУ ВШЭ, ТГУ, САФУ, СВФУ, УрФУ, МГУ по данным о выпускниках данных учебных заведений в социальной сети «ВКонтакте».

Представленная выше картосхема была построена на основе данных о выпускниках перечисленных учебных заведений, которые оставляют в своих профилях в социальной сети «ВКонтакте» информацию о месте жительства и учебы. На Рис. 2 каждому учебному заведению соответствует определенный цвет, который используется для отображения информации из каких городов поступают в этот университет. Например, МГУ соответствует розовый цвет. Можно заметить, что в МГУ поступают практически из всех представленных городов. Также на картосхеме указана информация о количестве профилей выпускников данного учебного заведения, которые были найдены в социальной сети.

Ограничения и риски использования цифровых следов в анализе миграционных потоков

Как и официальная статистика, данный метод сбора данных имеет ряд недостатков. Во-первых, не все пользователи социальной сети оставляют информацию о полученном образовании и миграциях, с ним связанных. Вместе с тем, существуют алгоритмы, позволяющие восстанавливать пропущенные данные через социальное окружение пользователя. Во-вторых, организация социальной сети не обязывает оставлять только достоверную информацию, и такие случаи необходимо уметь выявлять. В-третьих, возраст пользователей социальной сети смещен в сторону молодого населения.

Заключение

Анализ данных официальной статистики и цифровых следов пользователей социальных сетей позволит определить направления и масштабы миграции внутри страны в новом качественном ключе. Это новая возможность определить наиболее востребованные регионы и города, которые будут испытывать максимальный приток населения, выявить регионы в группе риска, которые будут испытывать кадровый голод. Предлагаемая методология позволяет дополнить официальную статистику качественными характеристиками (информацией, которые пользователи оставляют

в социальной сети: интересы, социальные связи, отношение к тем или иным вопросам и прочее). Дальнейшее конструирование геoinформационной системы [6], по аналогии с данными мониторинга востребованности вузовской сети [7], на базе полученных данных позволяет максимально упростить интерпретацию полученных результатов.

Список литературы:

1. Лешуков О. В., Евсеева Д. Г., Громов А. Д., Платонова Д. П. Оценка вклада региональных систем высшего образования в социально-экономическое развитие регионов России / Лешуков О. В., Евсеева Д. Г., Громов А. Д., Платонова Д. П. // Современная аналитика образования – 2017 – №3(11) – С. 1-31.
2. Абанкина И. В. Проблемы основной школы и система расселения России / Абанкина И. В. // Вопросы образования – 2005 – №2 – С. 94-111.
3. Кашницкий И. С., Мкртчян Н. В., Лешуков О. В. Межрегиональная миграция молодежи в России: комплексный анализ демографической статистики / Кашницкий И. С., Мкртчян Н. В., Лешуков О. В. // Вопросы образования – 2016 – №3 – С. 169-203.
4. Смирнов И. Б., Сивак Е. В., Козьмина Я. Я. В поисках утраченных профилей: достоверность данных «ВКонтакте» и их значение для исследований образования / Смирнов И. Б., Сивак Е. В., Козьмина Я. Я. // Вопросы образования – 2016. – №4 – С. 106-122.
5. «ВКонтакте» [Электронный ресурс]. URL: <https://vk.com> (дата обращения: 27.12.2018).
6. Рубцов В. А., Габдрахманов Н. К. Актуальность применения геопорталов в принятии управленческих решений (медико-демографические геoinформационные системы) / Рубцов В. А., Габдрахманов Н. К. // Экологический консалтинг – 2011 – №2(42) – С. 27-31.
7. CAO «Где учиться и где работать» [Электронный ресурс]. URL: <http://educational-migration.nextgis.com/resource/67/display?panel=layers&fbclid=IwAR2rHie2RZ1AdM1GAhpc6qJaHTQunxXdtND5HE5KGaWfYRtJ2N3Y5HN4csw> (дата обращения: 27.12.2018).

АВТОМАТИЗИРОВАННАЯ СИСТЕМА АНАЛИЗА ОРИГИНАЛЬНОСТИ КОДОВ ПРОГРАММ НА ЯЗЫКЕ PYTHON

М.А. Бубнова, Н.А. Мелех

*Национальный исследовательский университет
«Высшая школа экономики»,
департамент компьютерной инженерии
МИЭМ НИУ ВШЭ*

Аннотация

В работе приведены результаты исследования подходов к анализу оригинальности кодов программ лабораторных работ студентов. Разрабатываемая система входит в категорию систем поддержки принятия решения. Работа системы основывается на критериях, выделенных в ходе исследования, по которым работа считается списанной. Лицом, принимающим решение (ЛПР) выступает преподаватель, который, ознакомившись с выделенными системой как идентичные с заданным уровнем оригинальности работами, выносит окончательное решение.

Введение

Займствований в коде программных средств распространены аналогичным образом как и в статьях, рефератах и прочих работах. Проблема анализа оригинальности и выявления плагиата в программном коде остро стоит в текущее время, особенно перед лицом образовательных

учреждений. Однако при проверке частей программ необходимо использовать иные методы, отличные от методов анализа простых текстов, способных выявлять и не включать в плагиат такие вещи, характерные для программного кода, как принятые для языка программирования конструкции, выражения и структуры кода.

В работе приведены результаты, полученные при применении методов семантического и синтаксического анализа текста. Сравнение производилось на реальных работах студентов.

Решение

Разработанная система анализа оригинальности кодов программ представляет собой систему поддержки принятия решения для расчета процента оригинальности между всеми программными кодами, подающимися в качестве входных данных.

Система предназначена для работы с программным кодом в формате Jupyter Notebook (*.ipynb), т.е. работы на языке программирования Python.

Был выделен ряд критериев, которым должна соответствовать разработанная система в сравнении с существующими аналогами [1]:

- Доступность применения в повседневной работе без необходимости подключения к сети;
- Загрузка нескольких файлов одновременно;
- Возможность использовать систему без ожидания подтверждения регистрации от администраторов;
- Возможность выгрузки значений в числовом формате без графов;
- Возможность использования облачного хранилища для загрузки входных данных и выгрузки выходных данных. В качестве используемого сервиса выступает Dstorbox - для загрузки файлов и выгрузки отчетности.

Принципиально выделено два этапа для реализации задач, стоящих перед системой [2]. Первый этап заключается в предварительной обработке полученного текста для корректной работы некоторых методов, о которых речь пойдет далее. Вторым этапом является непосредственный анализ и сравнение программных кодов согласно выбранным алгоритмам.

Сравнение программных кодов происходит попарно для всех возможных комбинаций из набора элементов, подаваемого на вход системе. Анализ проводится с помощью методов, выделенных в ходе исследования:

- Строковый метод - тексты программ сравниваются дословно с учетом перекрытия текста. Данный метод аналогичен по применению как с программным кодом, так и с обычным текстом;
- Маркировочный - текст программы анализируется не полностью, выделяются лишь характерные для языка, на котором написаны сравниваемые программы, лексемы, представляющие собой наборы операторов (маркеров). Наборы составляются и указываются в системе, что позволяет корректировать степень детализации при проверке для различных наборов программ;
- Древовидный метод - метод, подразумевающий построение абстрактного дерева, описывающего структуру программы в соответствии с текстовым файлом кода [3]. Работа метода основывается в меньшей мере на синтаксисе и тексте, но в большей - на принципе работы программы. Между собой сравниваются построенные графы функционирования.

Был реализован алгоритм Вагнера-Фишера для расчета расстояния Левенштайна, определяющего схожесть двух строк. Выбранный алгоритм подходит для анализа полученных после предварительной обработки последовательностей лексем каждого программного кода. В результате коэффициент схожести двух программ вычисляется согласно уравнению: