Литературное программирование: от данных до презентации

Лаборатория экологии морского бентоса (гидробиологии) - работает со школьниками уже более полувека. На всем протяжении существования нашего объединения главной педагогической технологией было и остается привлечение учащихся к реальной научно-исследовательской деятельности. Нам, преподавателям, уже давно стало понятно, что в исследовательских проектах учащихся существует определенная динамика мотивации. На первых этапах выполнения проекта, которые включают планирование работы и сбор научного материала (участие в экспедициях, проведение наблюдений и экспериментов) школьники демонстрируют огромную заинтересованность и высокий градус мотивации. Однако на последующих этапах, через которые должен обязательно пройти проект (набивка базы данных, статистическая обработка данных, написание и оформление текста работы, подготовка публичных презентаций), степень искрометности значительно снижается. Некоторые школьники выдыхаются и иногда бросают проект. Самым сложным этапом является написание текстов. До недавнего времени основным средством преодоления этого кризиса были традиционные методы уговоров и мягкого давления. Однако в последнее десятилетие появилась возможность применять новую и весьма продуктивную технологию, которая, как не удивительно, связана с магией...

Один из активных популяризаторов науки, палеонтолог, д.б.н. К. Ю. Еськов в своей статье «Наш ответ Фукуяме» (http://lib.ru/PROZA/ESKOV\_K/pub\_fuj.txt) определил магию, как «способ непосредственно воздействовать при помощи информационных объектов на объекты вещественные». Это определение нисколько не противоречит имплицитному пониманию магии в литературе (см. сагу о Г. Поттере). Однако не трудно заметить, что именно такими воздействиями информации на материальный мир занимаются любые IT-технологии. Для современного человека, а уж тем более для школьника, такая форма существования магии, которая связана с управлением миром через программные коды (заклинания) посредством компьютеров (волшебная палочка), кажется абсолютно естественной. Дети быстро и без посторонней помощи овладевают многочисленными современными гаджетами, они, в отличие от их родителей, способны разобраться в программировании робота-пылесоса. Написание веб-страниц и управление интернет-сайтами для современного школьника уже не является чем-то из области задачи для высоколобого ученого.

IT-магия увлекательна и привлекательна. Оказалось, что именно эту привлекательность магии можно эффективно использовать для преодоления описанного выше кризиса мотивации в написании исследовательских работ.

Основой подхода для той модели работы с исследовательскими проектами учащихся, которая будет описана ниже, является два важнейших для академической науки тренда, которые распространились благодаря расцвету IT-магии. Во-первых, это представление об опрятных данных (tidy data) и культуре их преобразования и обработки (tidyverse). Во-вторых, это распространение идеологии воспроизводимого исследования (reproducable research). Обе эти составляющие лежат в основе техники, получившей название «литературное программирование». Эта техника позволяет организовать непрерывную цепочку от данных к их визуализации и статистической обработке и, далее, к финальному продукту, который имеет вид сверстанного текста или презентации. Все этапы этой цепочки обслуживаются программным кодом на простых и доступных для быстрого изучения языках. Не ставя перед собой задачи описания всех нюансов литературного программирования (это задача для полновесного курса, который мы предлагаем нашим учащимся), остановимся кратко на основных этапах указанной цепочки.

Для большинства людей, имеющих отношения к данным (от бухгалтеров до профессиональных ученых), набивка информации в те или иные электронныe таблицы (например, Excel) - занятие интуитивное. Документы, порожденные потоком наивного сознания, представляют для аналитиков, работающих с этими документами, немалую проблему и, самое главное, часто «вольное» обращение с первичными данными является источником ошибок. К сожалению, важнейшие принципы создания баз данных, пусть даже простых, в школьных курсах информатики не изучают в должной мере. Классический пример - это широчайшее распространение в таблицах, созданных в Excel, объединенных ячеек. Это категорически противоречит идее tidyverse! Поэтому на первом этапе освоения IT-магии мы специально учим наших ребят тому, как надо и не надо набивать первичные данные. Очень эффективной и увлекательной оказалось учебная задача создания базы данных со своими школьными оценками. На этом примере легко объяснять такие понятия как «поля базы данных», «запрос», «фильтрация» и т.п.

Когда первые шаги в магическом обучении пройдены и учащиеся смогли создать опрятную базу данных на основе материала своего исследовательского проекта, можно переходить ко второму уровню - это визуализация данных. Здесь незаменимым инструментом является язык статистического программирования «R». Этот скриптовый язык, на ряду с параллельно развивающимся языком «Python», стал золотым стандартом в обработке данных в биологии, социологии, психологии и прочих дисциплинах. Опыт показал, что освоение базовых навыков использования этого языка проходит за два-три занятия. Поскольку здесь школьникам раскрывается вся мощь IT-магии, когда коды программы приводят к извлечению из данных удивительных сведений, это резко повышает мотивацию. В частности, у учащихся огромный интерес вызывает теория грамматики графиков. Согласно этой теории, любой график - это не просто набор линий, а некоторый текст, который имеет свой синтаксис, пунктуацию и прочие атрибуты. Эта идея лежит в основе пакета «ggplot2», с помощью которого можно породить график любой сложности и эстетической красоты. Еще большую помощь язык R оказывает в освоении статистических методов обработки. С его помощью можно эффективнее и глубже объяснять труднейшие моменты статистики, такие как проверка гипотез, статистические тесты или линейные модели.

Наконец, третий уровень посвящения - это освоение собственно литературного программирования. Конечно, текст и смыслы, кроящиеся в нем, никто никогда не напишет в виде программы (даже современные технологии ИИ создают пока лишь жалкие подобия осмысленных сочинений). Это творческий процесс. Однако, его можно существенно облегчить и сделать более увлекательным, если по мере написания текста использовать языки описания разметки текста. Самым известным языком такого описания является LaTex, но опыт показал, что его осваивают лишь самые мотивированные школьники. Гораздо более простым языком, которым учащиеся овладевают очень быстро, оказался язык R Markdown. Этот язык позволяет комбинировать текст с кодом обработки данных (в том числе на языке R, а продвинутые учащиеся могут вставлять коды и на других языках). В результате получается очень строго организованная конструкция, где все текстовые фрагменты взаимодействуют с соответствующими рисунками и таблицами, которые получены, в свою очередь, на основе первичных данных проекта.

Технологии литературного программирования позволяют быстро и без потерь производить изменения в финальном тексте работы при изменении данных. Это особо важно для начинающих исследователей, так как часто бывает, что они, после критического осмысления своих материалов, что-то меняют (например, выбрасывают какие-то сомнительные измерения или добавляют новые результаты, которые не вошли в первую версию работы). Финальный продукт, который генерируется на основе кода R Markdown, имеет вид документа в формате HTML, Word или PDF. Столь же легко модифицировать этот документ в презентацию в формате Power Point или другие форматы (beamer, ioslides).

Вишенкой на торте произведения, полученного методами литературного программирования, оказывается построение библиографического описания процитированных в тексте литературных источников. Страшным сном, для всех, кто оформляет библиографические списки, является расстановка в описании источников разных значков (точки, тире, дефисы, косые черточки и т.д.). При этом порядок расстановки этих значков сильно варьирует в разных библиографических стилях. А уж если требуется оформление цитат с цифровыми сносками (например, по ГОСТ), то это вызывает головную боль на несколько дней. Все это делается быстро и безболезненно с помощью языка R Markdown. А главное, что это превращается в увлекательный квест, который учащиеся с удовольствием проходят.

Функциональная грамотность подразумевает способность человека использовать знания для решения задач в различных сферах. Несомненно, навыки приобретенные в результате освоения технологии литературного программирования в процессе работы над исследовательскими проектами, в значительной степени будут востребованы во многих областях взрослой жизни. Помимо творческой продукции, которая, возможно, составит основу профессии наших выпускников, им, несомненно, придется производить огромное количество менее осмысленных рутинных документов (отчеты, учебно-тематические планы, образовательные программы, расписания мероприятий, выпускные квалификационные работы и многое-многое другое). Эти документы, если они основаны на хорошо подготовленных данных, порождаются методами литературного программирования быстро и выглядят весьма солидно.

Завершая разговор следует отметить, что использование технологий IT-магии в формате приемов литературного программирования имеет очень много плюсов с точки зрения практикующего педагога. Прежде всего этот подход дисциплинируют ум учащихся. Во-вторых, технологии литературного программирования открывают широкие горизонты для взаимодействия педагога и ученика, появляется много осмысленных тем, которые интересны обоим (часто ученики быстрее и лучше понимаю некоторые новые вещи, например какой-нибудь неизученный ранее пакет, и могут сами обучать педагогов). В-третьих, любое программирование существенно способствует развитию навыков рациональной работы с информацией, позволяет критически осмыслять полученные данные. Это чрезвычайно важно в современных условиях, наполненных информационным шумом и разнообразными фейками. Ну, наконец, эти технологии основаны на открытых програмных продуктах. Открытое ПО может быть использовано только при постоянном контакте с окружающим миром (получение обновлений, чтение документации, справок и т.п.). Это позволяет избавиться от ощущения изоляции и приобщиться к современным трендам.

Однако во всей этой истории, есть один существенный минус, который, впрочем, может рассматриваться и как существенный плюс. Людям, которые только начинают приобщаться к IT-магии надо будет немного поработать над собой. Но это стоит потраченных усилий!