

ИНТЕРНЕТ-ПРИЛОЖЕНИЕ ДЛЯ ОБРАБОТКИ ДЕМОГРАФИЧЕСКИХ И МИГРАЦИОННЫХ ДАННЫХ

Л.В. Рудикова, А.И. Амбражевич, С.Л. Соболевский
Кафедра программного обеспечения интеллектуальных и компьютерных систем
Учреждение образования «Гродненский государственный университет имени Янки Купалы»
Лаборатория SENSEable City, Массачусетский Технологический Институт
Гродно, Республика Беларусь; Кембридж, Соединенные Штаты Америки
E-mail: rudikowa@gmail.com, mefor@tut.by, stanly@mit.edu

В статье приводятся основные теоретические и практические аспекты реализации Интернет-приложения для обработки данных, связанных с демографическими и миграционными показателями. Особое внимание уделяется алгоритму обработки миграционных данных для поиска сообществ в сети. Описывается общая архитектура разрабатываемого приложения.

I. АЛГОРИТМ АНАЛИЗА МИГРАЦИОННЫХ ДАННЫХ

ВВЕДЕНИЕ

В современном мире существует огромное количество статистической информации, систематизация и анализ которой позволит использовать ее для принятия каких-либо решений. Однако систем, работающих с демографическими и миграционными данными, крайне мало, что, весьма, удивительно, так как последствия миграции играют огромную роль в различных сферах человеческой жизни: политической, социальной, экономической, культурно-психологической и других. Несомненно, предлагаемая система будет востребована аналитиками, работающими в сферах деятельности, тесно связанных с демографией и миграциями.

Основной отличительной особенностью разрабатываемого Интернет-приложения является возможность анализа данных алгоритмом выделения сообществ, что позволит пользователю найти в них скрытые нетривиальные и полезные закономерности, позволяющих получить новые знания об исследуемых данных. Кроме того, предлагаемая система может сохранять демографические и миграционные данные, поэтому в системе имеется возможность предоставления указанных данных сторонним приложениям в виде соответствующего сервиса.

Итак, отметим основные возможности, которые поддерживаются системой: загрузка миграционных и демографических данных из внешнего источника в базу данных; наглядное отображение этих данных на интерактивной географической карте; возможность фильтрации данных по дате или иным критериям; возможность анализировать миграционные данные наиболее оптимальным алгоритмом нахождения сообществ, т.е. алгоритмом Лувейна; возможность просматривать результаты анализа на интерактивной географической карте возможность предоставлять демографические и миграционные данные сторонним приложениям.

Предметная область разрабатываемого Интернет-приложения непосредственно связана с демографией и направлена на решение поставленных перед ней задач.

Исследование населения посредством анализа основных демографических показателей является незаменимым и обязательным методом социальной диагностики. Так, своевременный и наиболее полный учет населения по демографическим показателям может предоставить полноценную картину демографической ситуации в стране, а, следовательно, определить направления дальнейшей социальной политики.

Для выявления новых знаний из исследуемых миграционных данных применяется алгоритм Лувейна [1] – алгоритм выявления сообществ в сети. Для использования указанного алгоритма преобразуем миграционные данные в взвешенный ориентированный граф, где вершинами будут географические регионы, а число вышедших из географического региона А в географический регион Б будет весом ребра.

Алгоритм состоит из двух шагов. На первом шаге происходит поиск «малых» сообществ путём оптимизации модульности на локальном уровне. На второй стадии узлы одного сообщества агрегируются и строится новая сеть большего масштаба, после чего эти стадии повторяются до тех пор, пока не будет достигнут максимальный уровень модульности. Таким образом, после каждого этапа отображаются сообщества всё большего масштаба (рис. 2).

Для оценки качества разбиения графа на сообщества, вводится понятие модулярности, описывающее, насколько при заданном разбиении графа на группы плотность внутрigrупповых связей больше плотности межgrупповых связей:

$$Q = 1/2m \sum_{i,j} [A_{ij} - k_i k_j / 2m] \delta(c_i, c_j)$$

где Q - модулярность, m - сумма весов всех ребер графа, $A_{i,j}$ - вес ребра соединяющего вершину i и j , k_i - сумма весов ребер выходящих из вершины i , $\delta(c_i, c_j) = 1$, если $c_i = c_j$, иначе 0.

II. ОБЩАЯ АРХИТЕКТУРА ИНТЕРНЕТ-ПРИЛОЖЕНИЯ

С учетом всех требований предметной области, а также возможности анализа и представления данных, для реализации Интернет-приложения выбрана многоуровневая архитектура [2], включающая совокупность следующих слоев: уровень хранения данных (состоит из базы данных), уровень доступа к данным (отвечает за предоставление выше лежащему слою контрактов для доступа к необходимым данным), уровень бизнес-логики (содержит в себе всю логику, связанную с манипуляцией над данными, реализацию алгоритмов и взаимодействие компонентов между собой внутри слоя), сервисный уровень (содержит в себе реализацию всех сервисов, с которыми могут взаимодействовать компоненты из уровня клиентов) и уровень представления (набор компонентов, которые ответственны за предоставление и отображение пользователю среды для работы с данными; в уровень представления также входят сторонние приложения, которые считаются потребителями сервиса).

Для реализации Интернет-приложения использованы следующие программные средства: технология .Net Framework компании Microsoft, язык C#, СУБД MS SQL SERVER, различные программные инструменты (Entity Framework, jQuery, amCharts), позволяющие решить поставленные задачи наиболее эффективным способом (рис. 1).

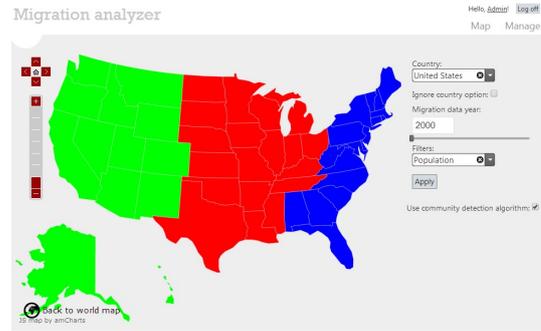


Рис. 1 – Алгоритм Лувейна

ЗАКЛЮЧЕНИЕ

Таким образом, предлагаемое Интернет-приложение для обработки демографических и миграционных, представляет собой многофункциональную систему, которая будет востребована людьми, заинтересованными в исследовании демографических показателей и анализе миграционных данных. Система предоставляет возможность наглядно вывести интересующие пользователя данные в удобном для него виде (на интерактивной географической карте), а также использовать к этим данным различные фильтрующие критерии. Важным аспектом данной системы является простота ее использования.

1. Louvain method: Finding communities in large networks – Mode of access: <https://sites.google.com/site/findcommunities/>. – Date of access: 28.03.2014.
2. Фаулер, М. Архитектура корпоративных программных приложений / М. Фаулер. – М.: Вильямс, 2007. – 544 с.

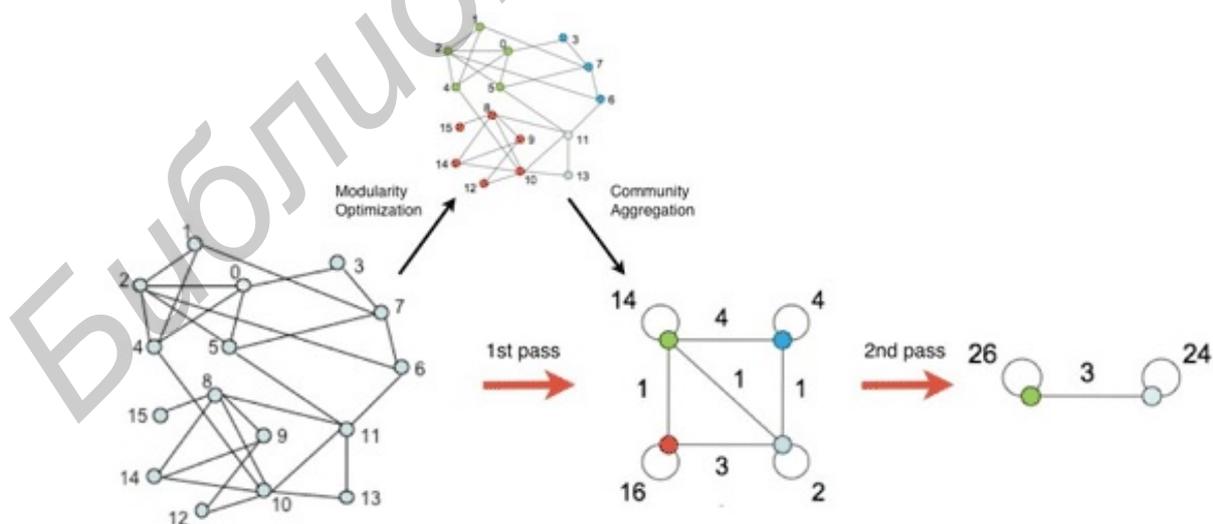


Рис. 2 – Алгоритм Лувейна