

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»
Факультет компьютерных наук
Департамент программной инженерии

СОГЛАСОВАНО

Доцент базовой кафедры «Системное
программирование» ИСП РАН

_____ Д.Ю. Турдаков
«___» _____ 2017 г.

УТВЕРЖДАЮ

Академический руководитель
образовательной программы
«Программная инженерия»

_____ В.В. Шилов
«___» _____ 2017 г.

**ИНСТРУМЕНТ КЛАСТЕРИЗАЦИИ И ВИЗУАЛИЗАЦИИ КЛАСТЕРОВ
НАУЧНЫХ СТАТЕЙ**

Программа и методика испытаний

ЛИСТ УТВЕРЖДЕНИЯ

RU.17701729. 507900-51-01-ЛУ

Подп. и дата	
Инв. № дубл.	
Взам. инв. №	
Подп. и дата	
Инв. № подл.	

Исполнитель:
студент группы БПИ133

_____ /Григорьев А.А. /
«___» _____ 2017 г.

УТВЕРЖДЕНО

RU.17701729. 507900-51-01-ЛУ

ИНСТРУМЕНТ КЛАСТЕРИЗАЦИИ И ВИЗУАЛИЗАЦИИ КЛАСТЕРОВ НАУЧНЫХ СТАТЕЙ

Программа и методика испытаний

Листов 16

2017

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729. 507900-51-01-ЛУ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

СОДЕРЖАНИЕ

ОБЪЕКТ ИСПЫТАНИЙ	68
ТРЕБОВАНИЯ К ПРОГРАММЕ	69
ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ	72
СРЕДСТВА И ПОРЯДОК ИСПЫТАНИЙ	73
МЕТОДЫ ИСПЫТАНИЙ	74

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729, 507900-51-01-ЛУ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ОБЪЕКТ ИСПЫТАНИЙ

Полное наименование программы – «Инструмент кластеризации и визуализации кластеров научных статей». Инструмент создаётся как компонент системы под названием Research Supporter и позволяет, разместив на исследовательской карте некоторое количество научных статей, автоматически разделить их на группы семантически близких документов и визуализировать полученные группы статей.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729, 507900-51-01-ЛУ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ЦЕЛЬ ИСПЫТАНИЙ

Целью испытаний является проверка правильности выполнения программой функций, перечисленных в разделе «Требования к программе».

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729, 507900-51-01-ЛУ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ТРЕБОВАНИЯ К ПРОГРАММЕ

1. Требования к функциональным характеристикам

Программа состоит из двух основных компонент: клиентской и серверной частей, между которыми должно быть налажено взаимодействие

1.1. Требования к серверной части

На серверной части должен быть реализован алгоритм кластеризации статей, разделяющий находящиеся в базе данных статьи, принадлежащие определённой **исследовательской карте** (research map), на группы таким образом, чтобы статьи в одной группе были семантически ближе по отношению друг к другу, чем по отношению к статьям из других групп.

Семантическая близость между статьями определяется как семантическая близость между текстами их заголовков и аннотаций.

Для определения семантической близости используется косинусное расстояние между векторными представлениями текстов с помощью алгоритма paragraph2vec [10].

Также должно быть реализовано взаимодействие с базой данных для получения статей и сохранения сгенерированных кластеров.

Каждый сгенерированный кластер должен быть представлен как структура, состоящая из собственного уникального по отношению ко всем сущностям в базе данных идентификатора и списка идентификаторов статей, относящихся к этому кластеру.

Должна быть возможность задавать количество кластеров, на которые будут разделены статьи.

1.2. Требование к взаимодействию клиентской и серверной частей

Взаимодействие между клиентской и серверной частями должно осуществляться посредством HTTP-запросов.

При получении GET-запроса от клиента, сервер должен ответить сообщением в формате JSON, содержащим список сгенерированных кластеров с их уникальными идентификаторами и идентификаторами статей, относящихся к ним.

1.3. Требования к клиентской части

Клиентская часть должна быть реализована в виде веб-приложения, запускающегося в браузере и представлена в виде интерактивной карты с расположенным на ней графом цитирования. В графе цитирования статьи являются узлами графа, а ребрами представляется факт наличия одной статьи в списке источников другой. На карте же узлы графа цитирования отображаются в виде

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.507900-51-01-ЛУ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

прямоугольных элементов с текстовой информацией о статье, а рёбра -- стрелками от цитирующей статьи к цитируемой.

Веб-приложение должно предоставлять следующие возможности:

- разделить на кластеры все статьи на карте;
- разделить на кластеры выбранные статьи;
- разделить на новые кластеры статьи, помещённые в выбранные кластеры;
- удалить выбранные кластеры;
- удалить все кластеры;

Также при разделении статей на кластеры пользователь должен иметь возможность задать количество получаемых кластеров.

Каждый кластер должен быть представлен в виде замкнутого контура, внутри которого располагаются все узлы, представляющие статьи, относящиеся к данному кластеру.

Данные контуры должны автоматически перерисовываться в результате перемещения узлов статей по исследовательской карте.

Кроме того, визуальные отображения кластеров должны взаимодействовать между собой таким образом, чтобы минимизировать площадь пересечения их внутреннего пространства.

2. Требования к надежности

2.1. Требования к обеспечению надежного (устойчивого) функционирования программы

Пользователю, работающему с программой через веб-браузер должен быть предоставлен непрерывный доступ к веб-приложению, расположенному по определённому url-адресу. Веб-сервис не должен непредвиденно прерывать свою работу.

2.2. Время восстановления после отказа

В случае отказа работы серверной части и последующей недоступности веб-приложения, время восстановления не должно превышать одних рабочих суток.

2.3. Отказы из-за некорректных действий оператора

После запуска программы на сервере, отказ программы вследствие некорректных действий оператора должен быть исключён. В том числе, должна быть исключена возможность непреднамеренного выключения программы не связанного с техническими неполадками сервера.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729. 507900-51-01-ЛУ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ

Состав программной документации

- 1) «Инструмент кластеризации и визуализации кластеров научных статей». Техническое задание (ГОСТ 19.201-78);
- 2) «Инструмент кластеризации и визуализации кластеров научных статей». Программа и методика испытаний (ГОСТ 19.301-78);
- 3) «Инструмент кластеризации и визуализации кластеров научных статей». Руководство оператора (ГОСТ 19.505-79);
- 4) «Инструмент кластеризации и визуализации кластеров научных статей». Код программы (ГОСТ 19.401-78).

Специальные требования к программной документации

- 1) Все документы к программе должны быть выполнены в соответствии с ГОСТ 19.106-78 [6] и ГОСТ к этому виду документа (см. п. 5.1.).

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729, 507900-51-01-ЛУ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

СРЕДСТВА И ПОРЯДОК ИСПЫТАНИЙ

Технические средства

Для проведения испытаний использовались следующие два устройства.
Компьютер со следующими характеристиками:

- 64-разрядный (x64) процессор с тактовой частотой 1.3 гигагерц (ггц);
- 4 гб оперативной памяти (озу);
- 128 гб пространства на жестком диске;
- устройства ввода (клавиатура, мышь)
- устройство отображения (монитор)
- доступ к сети Интернет
-

Программные средства

Испытания на компьютере осуществляются на операционной системе Ubuntu 16.04 с помощью браузера Chrome. Веб-сервис был запущен локально на используемом компьютере.

Порядок проведения испытаний

Для начала проведения испытаний необходимо запустить на компьютере систему управления базами данных Apache Cassandra, а затем запустить веб-сервис, зайдя в корневую папку программы и вызвав команду “./activator run”.

После того как веб-сервис запущен можно зайти на страницу исследовательской карты (в нашем случае расположенную по адресу localhost:9000/text-clustering) и начать проведение испытаний.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729, 507900-51-01-ЛУ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

МЕТОДЫ ИСПЫТАНИЙ

В процессе испытаний был протестирован функционал веб-сервиса касающийся разделения набора расположенных на исследовательской карте статей на кластеры.

Изначально на карте были расположены 9 научных статей найденных с помощью сервиса Google Scholar по запросу “Petri nets” (рис.1).

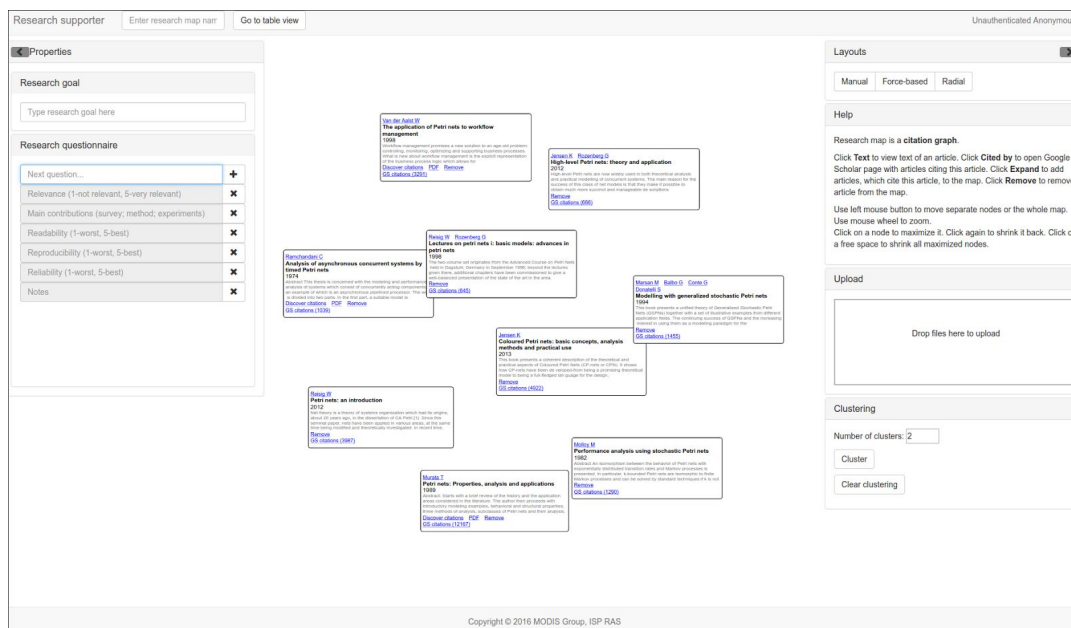


Рисунок 1

На панели кластеризации выбирается количество кластеров, на которые следует разделить статьи (рис. 2).

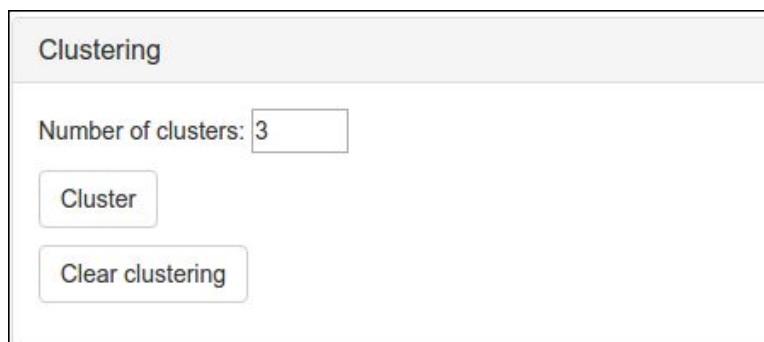


Рисунок 2

После нажатия на кнопку “Cluster”, статьи были визуально разделяются на выбранной количество кластеров. Каждый кластер представлен в виде замкнутого контура, внутри которого располагаются все узлы, представляющие статьи, относящиеся к данному кластеру (рис. 3). При перемещении узлов статей, кластеры меняют свою форму так, чтобы по-прежнему включать в себя все соответствующие статьи.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729, 507900-51-01-ЛУ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

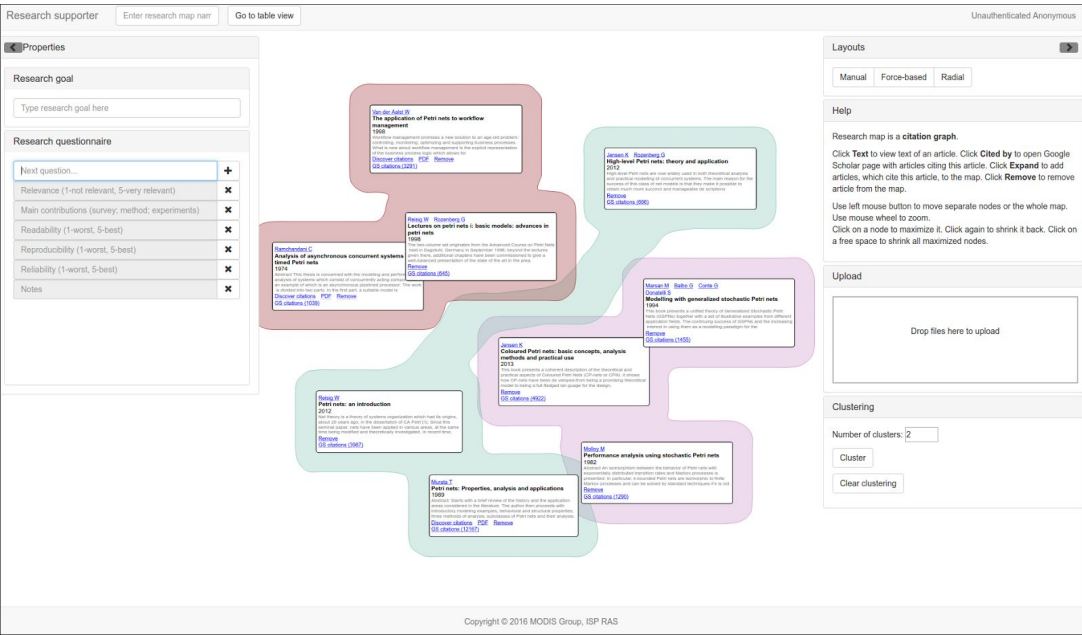


Рисунок 3

При нажатии левой кнопкой мыши на пространство одного из кластеров, он должен выделиться, а панель кластеризации изменить свой вид как показано на рисунке 4.



Рисунок 4

При нажатии на кнопку “Delete cluster”, контур выбранного кластера должен удалиться с исследовательской карты (рис. 5).

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729, 507900-51-01-ЛУ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

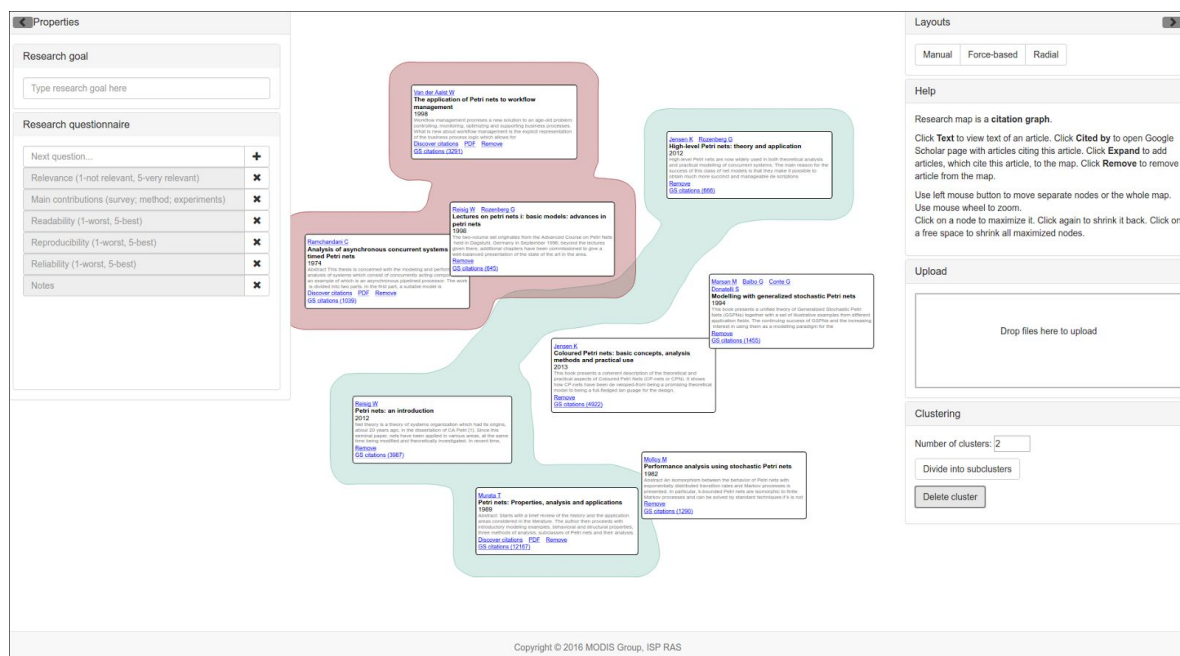


Рисунок 5

При нажатии на кнопку “Clear clustering” с карты должны удаляться все кластеры (рис. 6).

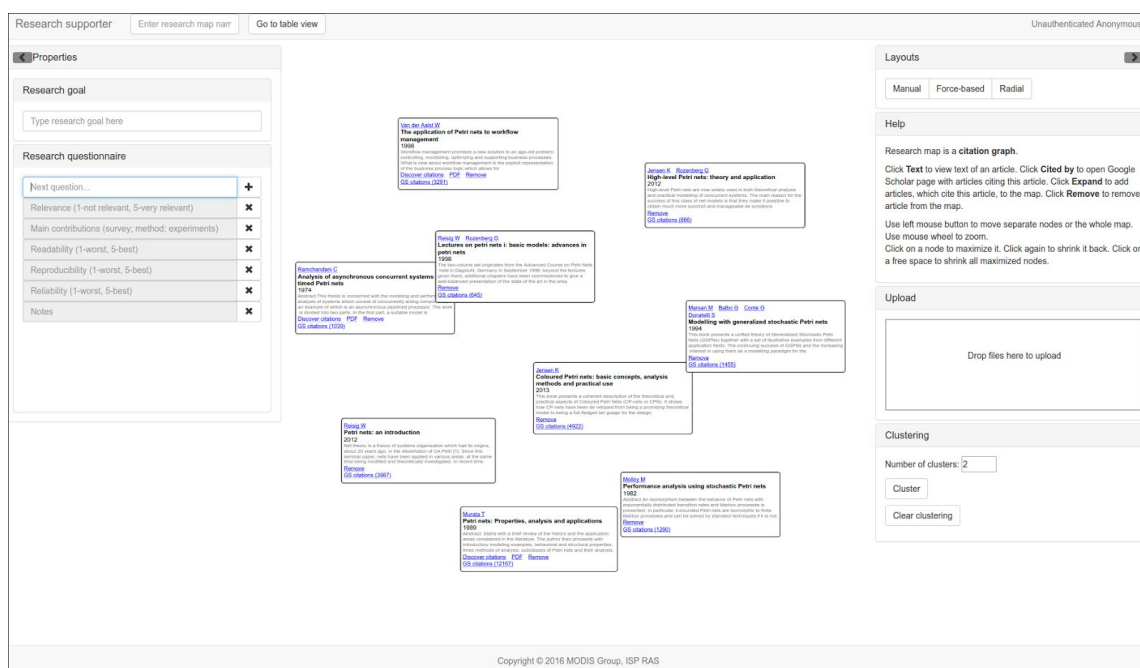


Рисунок 6

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729, 507900-51-01-Л/У				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

При выделении нескольких статей и нажатии на кнопку "Cluster", на кластеры должны разделяться только выбранные статьи (рис. 7).

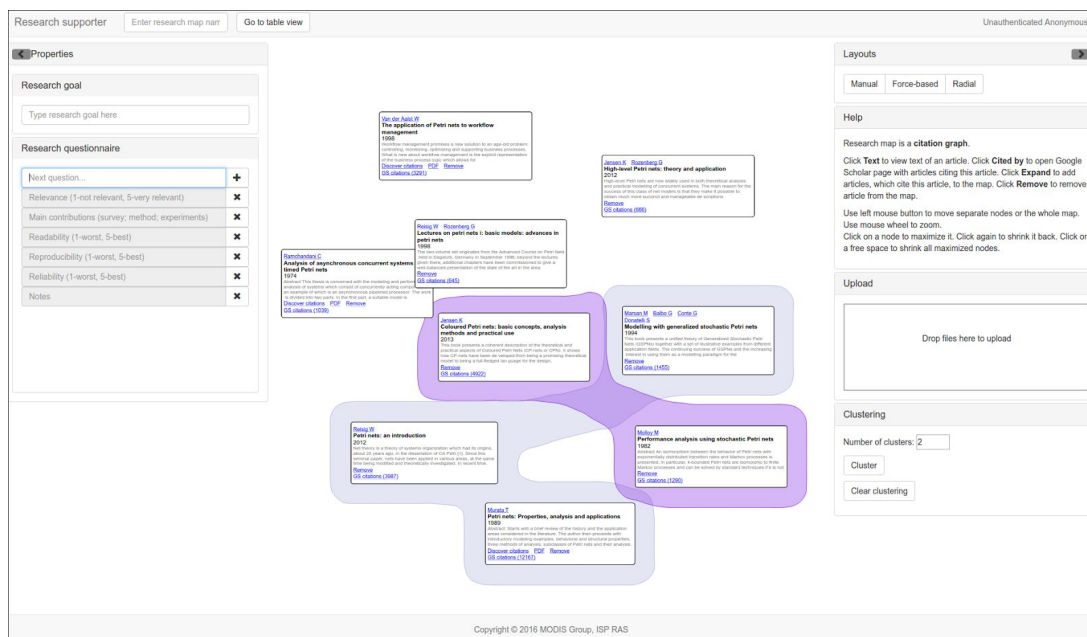


Рисунок 7

При выделении одного или нескольких кластеров, на панели кластеризации появляется кнопка "Re-cluster selected" (рис. 8).



Рисунок 8

При нажатии на неё, выбранные кластеры удаляются, и статьи, которые к ним относились, разделяются на новые кластеры (рис.9).

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729, 507900-51-01-Л/У				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

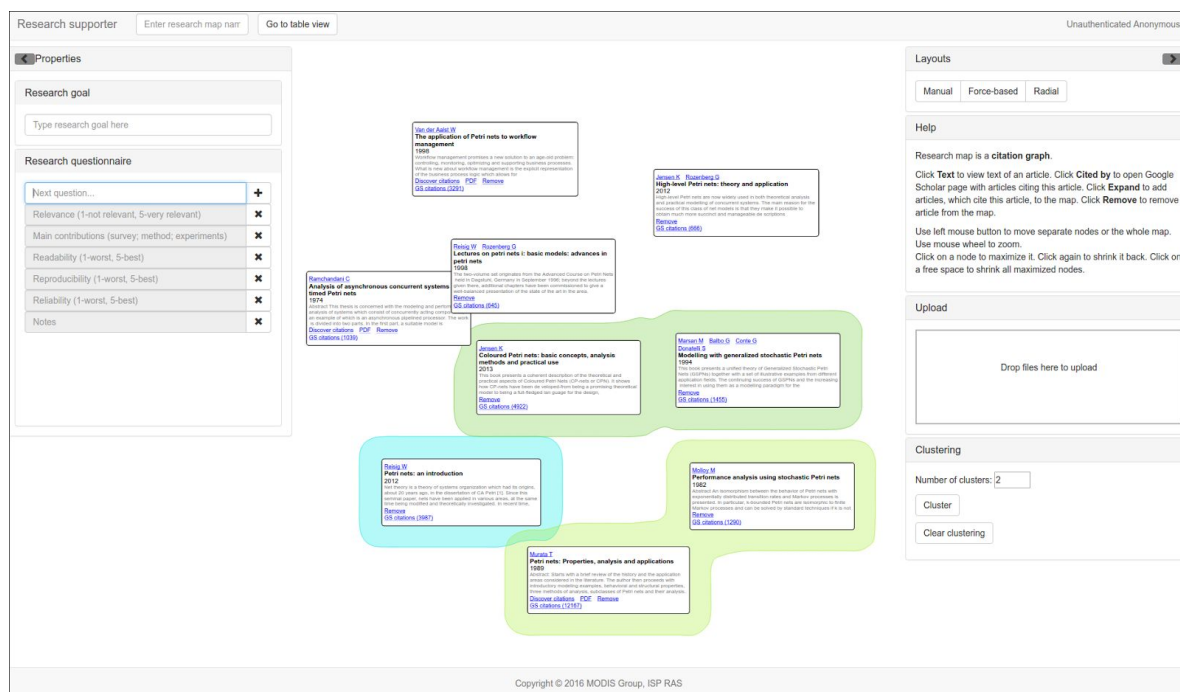


Рисунок 9

За все время использования системы, сбоев замечено не было.
Все испытания были проведены. В их результате нарушения требований к программе выявлено не было.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729, 507900-51-01-ЛУ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ПРИЛОЖЕНИЕ 1

ГЛОССАРИЙ

Понятие	Определение
Исследовательская карта	Интерактивное svg-изображение, на котором отображается граф цитирования и кластеры статей.
Научная статья	Статья опубликованная в научном журнале.
Исследователь	Научный работник проводящий исследования.
Список источников	Список ресурсов, на которые ссылается данная научная статья, расположенный, как правило, в конце статьи.
Контур	Непрерывная замкнутая кривая линия.
Панель кластеризации	Элемент интерфейса веб-приложения, на котором расположены элементы управления кластеризации.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729, 507900-51-01-ЛУ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

Лист регистрации изменений									
Номера листов (страниц)					Всего листов (страниц в докум.)	№ докумен та	Входящий № сопроводительн о- го докум. и дата	Подл.	Дата
Изм.	Изме нен- ных	Замене н- ных	нов ых	аннули рованн ых					

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729, 507900-51-01-ЛУ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата