Tesis de Maestría. Universidad ITMO, San Petersburgo, Rusia.

Разработка модели поиска в информационной семантической системе.

López-Pablos, Rodrigo.

Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

“САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ”

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

«РАЗРАБОТКА МОДЕЛИ ПОИСКА В ИНФОРМАЦИОННОЙ СЕМАНТИЧЕСКОЙ СИСТЕМЕ»

Автор Лопес-Паблос, Родриго _____
(Фамилия, Имя, Отчество) (Подпись)

Направление подготовки (специальность) 09.04.01Математические модели и компьютерное моделирование

Квалификация магистр
(бакалавр, инженер, магистр)

Руководитель Демин А.В., профессор, д.т.н. _____
(Фамилия, И., О., ученое звание, степень) (Подпись)

К защите допустить

Зав. кафедрой Муромцев Д.И., доцент, к.т.н. _____
(Фамилия, И., О., ученое звание, степень) (Подпись)

“___”_____ 20 ____г.

Санкт-Петербург, 2019 г.

Студент Лопес-Паблос, Родриго Группа Р4215 Кафедра ИПМ Факультет ПИиКТ
(ФИО)

Направленность (профиль), специализация Математические модели и компьютерное моделирование

Консультант(ы):

а) _____ _____
(Фамилия, И., О., ученое звание, степень) (Подпись)

б) _____ _____
(Фамилия, И., О., ученое звание, степень) (Подпись)

Квалификационная работа выполнена с оценкой _____

Дата защиты “11” Июня 2019 г.

Секретарь ГЭК _____

Листов хранения _____

Демонстрационных материалов/Чертежей хранения _____

ОГЛАВЛЕНИЕ

СПИСОК СОКРАЩЕНИЙ

БД – База данных

БСД – Байесовская сеть доверья

ОДР – Обучения дерева решений

ГВМ – Графовая вероятностная модель

ИАД – Интеллектуальный анализ данных

ИНС – Искусственные нейронные сети

ИСС – Информационная семантическая система

ПО – Программное обеспечение

СКК – Самоорганизующиеся карты Кохонена

ЭИ – Методы эксплуатаций информации

ВВЕДЕНИЕ

Актуальность темы исследования.

В настоящее время ни одна сфера деятельности человека не обходится без использования информационных технологий, что является основной характеристикой информационной эпохи. Информация и данные, которые ее составляют, часто содержат скрытые знания, глубину, неоднородности и значимость, выявление которой выходит за рамки простой обработки. Осуществление выявления и обработки этой неявной, скрытой информации является актуальным вопросом, поскольку это позволяет анализировать и принимать решения в рамках сложных реальных систем.

«Информационная семантическая система» (ИСС) – это информационная система, предназначенная для сбора, обработки и представления информации. Информационная семантическая система как правило реализуется на физическом и логическом уровне обработки информации [5, 7].

История науки и инженерии показывает, что обработка информация является единственным способом открытия новых знаний, а также наиболее эффективным методом принятия решений в любой организации или моделирования любой системы.

Чтобы смоделировать любую сложную систему, сначала необходимо проанализировать «скрытую» информацию, присутствующую в доступных данных, а также оценить уровень и глубину сложности анализируемой системы.

В данной работе представлена модель информационно-поисковой системы, использующая интеллектуальный анализ данных (ИАД) в качестве инструмента преобразования данных в исходном виде в семантическую информационную систему, содержащую и представляющую новые полезные знания, извлеченные из исходных данных.

В общей теории систем есть понятие сложной системы, которая рассматривает систему как совокупность взаимосвязанных элементов, образующих подсистемы и другие части системы. Информационная семантическая система (ИСС) отличается от сложной системы принципиально тем, что состоит из качественно разнородных элементов. Этими элементами являются семантические информационные единицы [5].

Данная выпускная квалификационная работа рассматривает следующие вопросы исследования:

- Могут ли средства интеллектуального анализа данных помочь ускорить, упростить или иначе усовершенствовать процесс извлечения и анализа «скрытой» информации, содержащейся в некотором наборе данных?
- Как можно смоделировать «неявное» знание?
- Можно ли моделировать информационный состав сложных неоднородных данных более точной характеристикой, чем числом бит?
- Какие методы и комбинации методов интеллектуального анализа данных могут быть эффективно реализованы для открытия новых знаний?
- Возможна ли проверка данных методов на реальных базах данных?

- Какие новые знания получится извлечь при помощи данных методов в конкретном прикладном случае?

Для выявления неявных знаний, например скрытой связи между параметрами, применяют специальные методы анализа, к числу которых относится коррелятивный анализ. В данной выпускной квалификационной работе также предлагается использовать комбинации методов интеллектуального анализа данных.

Также в данной выпускной квалификационной работе рассматривается применение комбинаций методов интеллектуального анализа данных в качестве инструмента для обнаружения скрытых правил в данных и их взвешивания, а также внутренних взаимосвязей в массивах анализируемых данных. Сочетание и комбинирование приемов ИАД можно назвать процессом эксплуатации информации.

Рассматриваемые процессы ЭИ требует значительных вычислительных ресурсов, и, используя аналитические инструменты и синтез, превращающие информацию в знания, стремятся генерировать знания, способствующие принятию решений. (Бритос и Гарсия Мартинес [10]) Такие технические системы применимы для построения модели поисково-информационной системы.

Степень теоретической разработанности темы. Так как само понятие эксплуатации информаций относительно ново, [10] теоретических исследований в области эмпирического междисциплинарного подхода к информации и информационной философии за пределами классического количественного подхода сравнительно мало. Однако были найдены труды, составляющие информационную основу исследования.

Цель и задачи исследования. Повысить качество и эффективность процесса выявления «скрытых» знаний и закономерностей в больших объемах данных посредством построения универсальной информационно-семантической системы, позволяющей тестировать и комбинировать методы поиска информации и знаний в любой заданной базе данных.

Для достижения данной цели поставлены следующие задачи:

I. Провести аналитический обзор методов представления информации и средств обработки данных.
II. Провести обзор и анализ систем эксплуатации информации для выявления подходов к проектированию такой системы.
III. Провести анализ алгоритмов интеллектуального анализа данных и их функциональных возможностей для поиска информации в семантической системе.
IV. Спроектировать универсальную информационно-семантическую систему и произвести эксперименты с реальными и открытыми базами данных.

Область исследования. Проведенное исследование принципов процесса эксплуатации информации с использованием методов интеллектуального анализа данных, процесса эксплуатации информаций и последующей эмпирической обработки соответствуют специальности «Математический и вычислительное моделирование», а содержание диссертации - техническому заданию.

Объект и предмет исследования. В качестве объекта исследования выбраны методы интеллектуального анализа данных для выявления «скрытых» знаний в наборе данных. Предметом исследования являются возможные варианты оптимизации систем анализа данных и

связанных с ними процессов, непосредственно влияющих на разработку таких систем.

Теоретическая и методологическая основа исследования. Методологической основой исследования является эвристический и эмпирический подход в отношении комбинаций методов интеллектуального анализа данных. Теоретической основой данной выпускной квалификационной работы являются библиографические источники, основным из которых стала статья Гарсия-Мартинес и Вритос «Предложение процессов эксплуатации информации» [9].

Информационная база исследования. База данных (БД) аффидавитов законодателей аргентинской думы. Данная открытая база данных содержит 1019 экземпляров недвижимости, заявленных аргентинским законодательством с 2003 по 2012 годы.

Научная новизна исследования. Исследованы ранее не изученные комбинации алгоритмических процессов в области интеллектуального анализа данных. Полученный результат позволяет достичь нового подхода к анализу баз данных, сократить время извлечения полезной информации из наборов данных, и впоследствии строить модели принятия решений на основе данных любого рода.

Практическая значимость исследования. Разработанный метод может быть использован для обработки любых наборов данных с целью выявления неявных знаний и преобразования данных в информацию с семантическим значением.

Апробация результатов исследования. Разработанная информационно-семантическая система была опубликована в сборнике трудов XLVIII международной научной конференции ежегодного собрания всеаргентинской ассоциации политической экономики в рамках доклада на тему «Элементы информационной эксплуатации

применительно к налоговому расследованию бюджета» (г. Росарио, Аргентина, 2013г. [11]).

1. ФИЛОСОФИЯ ИНФОРМАЦИИ, ПРОБЛЕМЫ МОДЕЛИРОВАНИЯ ИНФОРМАЦИИ

1.1 Проблемы философии информации

Проблемы, связанные с информацией, были в основном связаны с ее числовым качеством, а не с семантической глубиной. В то время как данные можно считать представительными для сложной системы, семантические информационные системы обладают знаниями, которые можно применять для принятия решений, поскольку они обладают неявными знаниями и помогают общему пониманию системы.

В философии выработалось особое понимание проблемы знания, оно определяется главной задачей, которую всегда решали философы: понять отношение человека к миру.

Семантическая теория информации, по которой главным является содержательность информации, (а не информационная емкость). Главным критерием наличия семантики (содержательности) в информации определяет истинность информации с позиций эпистемологии [5].

Данная выпускная квалификационная работа пытается приблизиться к решению некоторых открытых проблем в философии информации, поднятой Лучиано Флориди [9] чтобы решить какую, простое измерение битов или кубитов (в случае использования квантовых компьютеров) абсолютно бесполезно. Ниже, некоторые из открытых проблем, которые будут рассматриваться в этом дипломе.

Таблица 1.1 – Открытые проблемы в информационной философии Флориди – Проблемы № 4,5,7,11 [9].

Проблема обоснования данных	Как могут данные нести смысл?
Проблема правдивости	Как могут осмысленные данные нести истинное значение?
Проблема информационной семантики	Может ли информация эксплицировать значения?
Проблема семантического взгляда на науку	Можно ли свести науку к информационному моделированию?

Как видно из таблицы 1.1 для решения (или, по крайней мере, постановки упомянутых) проблем теория информации Шеннона оказываются совершенно бесполезной.

В таблице 1.2 приведен простой пример бесполезности количественного подхода для нахождения семантического значения в данных.

Таблица 1.2 – Пример количественной оценки информации, явно подчеркивающий наличие семантической информации, которую невозможно оценить прямым методом подсчета количества информации.

Данные	Количество знаков и символов
Казнить нельзя, помиловать	24 знака и 2 пробела
Казнить, нельзя помиловать	24 знака и 2 пробела
Нить, ватьнепомльзаказило	24 знака и 2 пробела

Очевидно, первая и вторая фразы противоположны по смыслу, а третья бессмысленна, хотя по теории информации Шеннона все фразы

эквивалентны, следовательно, мера Хартли и Шеннона не определяет количество информации по её смысловой нагрузке.

Одной из задач любой науки является получение и формирование информационных ресурсов в предметной области данной науки [8]. Современные информационные ресурсы включают различные компоненты: данные, информацию, описания, базы данных, знания и технологические системы [4, 5]. В информационном поле присутствует такой феномен как «неявное» знание, такое обозначает наличие отношений и характеристик в природе информации.

Такое отношение известно как иерархия информации, которая может быть представлена следующим образом на Рисунке номер один.

Рис. 1. Иерархия информации

На Рисунке 1 отображается каждый уровень добавляет определённые свойства к предыдущему уровню независимо от области науки, что во многих науках используются разные информации - оттуда междисциплинарный характер информации философии. Также, отношение «данные – информация – знания» [8] служит основой получения знаний. И его же служит основой извлечения знаний для формирования информационных ресурсов [4]».

В основании находится уровень «данных», которые могут быть взяты в качестве «сырья» для всех явлений во вселенной, «Информация» добавляет контекст и «Знание» добавляет «как» (механизм использования). Знание и есть то, что связывает человека с миром, говорит ему о реальности [7].

Гипотетически, даже за пределами «Знание» будет уровень «Мудрость» добавляет «когда» (условия использования), которые в этой работе мы включаем в иерархию «Знание» [1].

Именно на уровне «знания» и «мудрости» лежит такое «неявное» знание, где находится система семантических единиц, несущая полезную информацию. Информационная семантическая система (ИСС) представляет собой структурную систему связанных семантических единиц. ИСС реализована на логическом уровне описания информации [7].

Тем не менее, несмотря на свою логичность, самая большая проблема в моделировании знаний и полезной информации заключается в характеристиках и абстрактном аспекте этих «неявное» знаний. Проблема заключается в том, как смоделировать эту уникальную логичность, по крайней мере, в ее представлении, но проложив путь к открытию новых полезных знаний, как обсуждалось в следующем разделе.

1.2 Модель семантического поиска

Информационная семантическая система отличается от сложной системы принципиально тем, что состоит из качественно разнородных элементов [5]. Этими элементами являются семантические информационные единицы. Эта единица имеет полезность и позволяет решить определенную проблему относительно момента и определенной ситуации, что делает его очень трудно быть представленным и измеренным.

Так можно ли смоделировать абстрактные знания? По крайней мере изначально, если предложение основано на общем определении информации (General Definition of Information), подход Винера, а не Шеннона, развивает Лучиано Флориди, когда делает попытку ввести [7, 5].

Так как единица битов и кубитов не может рассматриваться как единица анализа для представления системы семантической информации, единица информации определяется с возможностью обладать смыслом сама по себе.

Определяя эту семантическую единицу как «σ», является экземпляром информации, понимаемой как смысловое содержание, если и только если соблюдаются следующие общие определения информации (ООИ):

(ООИ.1) «σ» состоит из одного или нескольких данных;
(ООИ.2) данные в «σ» хорошо сформированы;
(ООИ.3) хорошо сформированные данные в σ являются значимыми;

Проблема попыток смоделировать информацию, превышающую количество битов, сталкивается с семантической субъективностью, полезностью и значением информации в ее человеческом применении.

Следуя математической поисковой модели Демина [2] Формулировка предложения адаптирована для поиска значимой семантической информации в базе данных следующим образом.

$$ICC = f(V(\Omega(\sigma)); \tau;\{a\}) \quad (1)$$

Где ICC представляю семантическую информационную систему (ИСС), способную решить проблему или принять решение т. к. преобразование данных в информацию и знания.

Поле анализа (σ) – часть области поиска, где осуществляется непосредственное обнаружение объекта поиска, т. к. ИСС. Область поиска данное пространство, в которой осуществляется поиск семантической система объекта (значимый смысловой объект).

Ω(σ) – Энергия, способность и подготовка наблюдателя к анализу данных.

V(Ω(σ)) – Объем и сложность данных обнаружения Ω(σ).

τ – Необходимое время анализа Ω(σ) для принятия решения о наличии или отсутствии полезной информации в используемых данных.

a – Несоответствия, потеря данных, мисингс, ошибка выборки и.т.д.

До тех пор, пока объем данных и их сложность V(Ω(σ)) будут больше, будет время наблюдения и созерцания τ, чтобы достичь I.

Однако, чем больше возможностей и подготовки наблюдателя, тем меньше времени и энергии потребуется для достижения значения I.

Логически добавить инструменты интеллектуального анализа данных (ИАД) и методы эксплуатации информации (ЭИ – сочетание методов ИАД) иметь эффект, сравнимый с увеличением способности наблюдателя σ.

Поисковое усилие – усилие, затрачиваемое на единицу объема анализируемой информации (например, время, затрачиваемое на просмотр поля анализа).

Суммарное поисковое усилие, приходящиеся на анализ всей данных; функция распределения поисковых усилий по обследуемой базе данных.

Обеспечение поискового ИСС, которое обеспечивает попадание ИСС в поле анализа при известных ограничениях на область, наличие проблемы и время поиска – этап эвристической обнаружения.

Естественно, смысловое значение, заложенное в семантической информационной системе «σ» не может быть определено количественно значение, а то, которое может быть достигнуто только с помощью индуктивного и эвристического подхода.

Основой взаимодействия является информационное взаимодействие. Оно может быть пассивным (созерцание, наблюдение). Оно может быть активным (измерение, воздействие, эксперимент). Пассивное можно обозначить как информирование, активное как собственно взаимодействие. В информационном поле присутствует такой феномен как «неявное» знание [7].

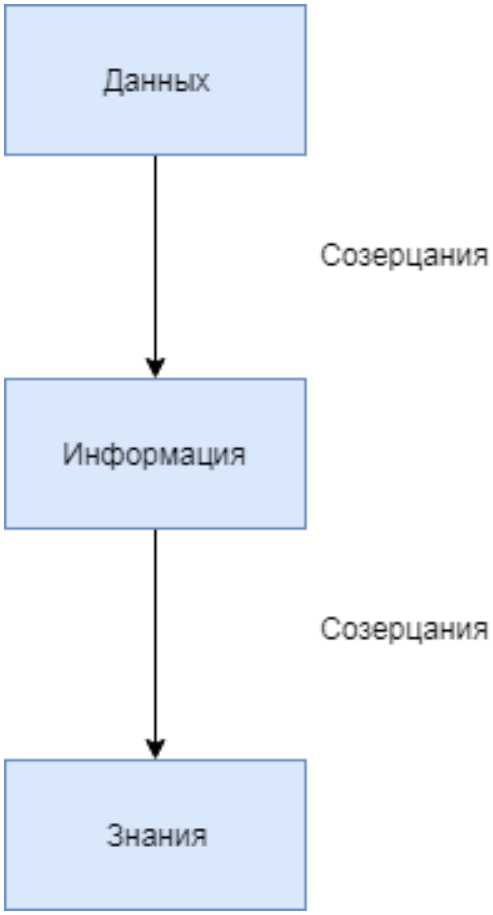

Рис. 2. Стандартный процесс открытия знаний

Рисунок 2 соответствует к «стандартному» процессу поиска полезной информации. Типичное исследование данных состоит из следующих пунктов:

I. Построение проекта исследования анализа данных.
II. Созерцание, наблюдение за данными, статический анализ данных.
III. Изучение результатов.
IV. Приход к новой информации и знаниям.

Стандартный подход все еще актуален и полезен. Однако единственному изучению данных и применению традиционных статистических средств, поиск информации и знаний может привести к

увеличению человеческого времени, усилий и энергии. Особенно в текущем контексте информационной перегрузки и помех.

Для выявления неявных знаний, например скрытой связи между параметрами, применяют специальные методы анализа, к числу которых относится коррелятивный анализ [6]. Среди этих методов есть множество алгоритмов интеллектуального анализа данных, которые могут быть использованы для поиска таких единиц.

Однако использование инструментов классической статистики и математической регрессии не исключает дополнительного использования методов интеллектуального анализа данных. Зачастую, методы интеллектуального анализа данных оказываются дополняющими методы классической статистики и математической регрессии, становясь частью более сложной, составной системы обработки данных и выявления информации.

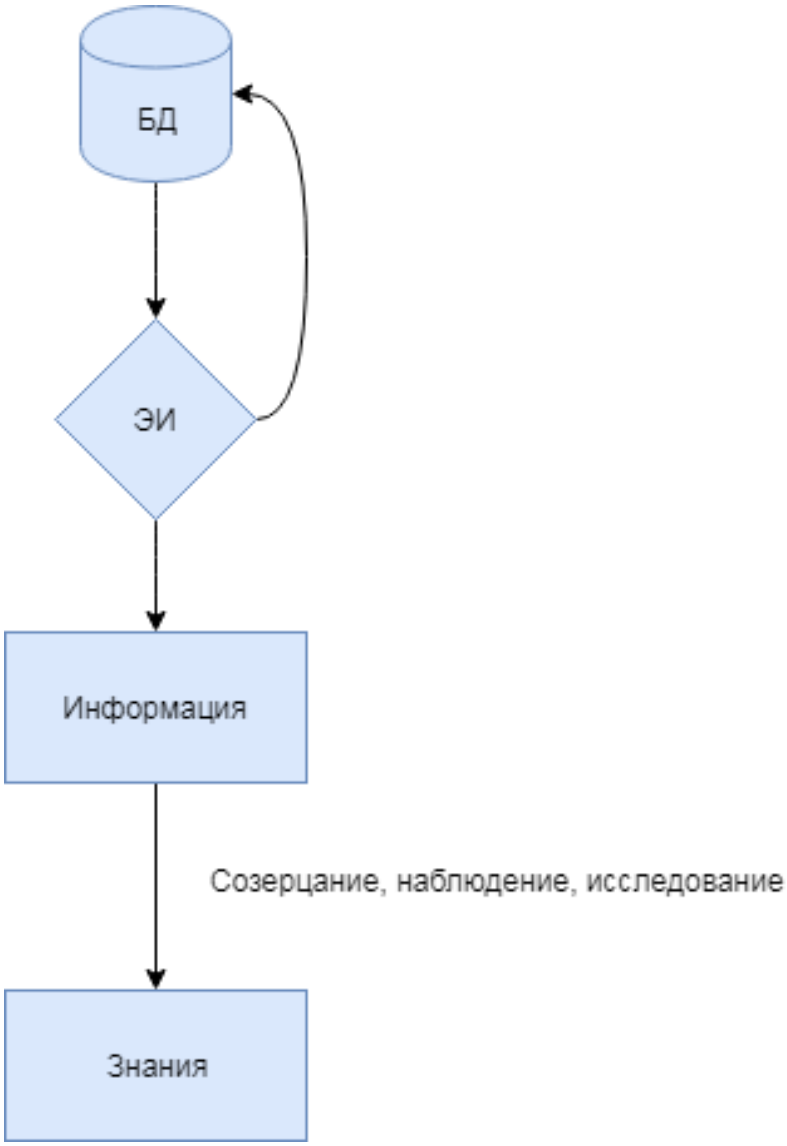

Рис. 3. Система поиска ИСС с поддержкой методов эксплуатации информации (ЭИ)

Схема с использованием разрабатываемого методом ЭИ будет выглядеть иначе (Рисунок 3). Включение использования эксплуатации информации для анализа данных в поисковой системе знание достигаются следующие улучшение:

Уделено более совершенному процессу «ЭИ», но в то же время желательно, чтобы найденное решение было возможно применить.

I. Построение проекта исследования анализа данных.
II. Применение процессов ЭИ.
III. Изучение результатов.
IV. Приход к новой информации и знаниям.

Принимать обоснованные управленческие решения по фактическим данным, тенденциям и закономерностям.

Согласно уравнению (1) включение методов ЭИ позволит достичь семантических единиц полезной информации с меньшими усилиями и меньшим временем ($< \tau$).

Более того, позволяют обнаруживать и исправлять несоответствия и потерю данных ($< a$), независимо от объема, сложности $V(\Omega(\sigma))$ и энергии $\Omega(\sigma)$, которая используется для обнаружения и понимания существования некоторой логической ICC.

Это так, учитывая способность вычислительного инструмента поддерживать и умножать ограниченные биологические возможности человека.

2. МЕТОДОЛОГИЯ И АЛГОРИТМ ПОСТРОЕНИЯ МОДЕЛИ ПОИСКА В ИНФОРМАЦИОННОЙ СЕМАНТИЧЕСКОЙ СИСТЕМЕ

2.1 Обзор и анализ систем эксплуатации информации (ЭИ)

Эксплуатация информации (EI, испан. Explotación de la Información) – это «подразделение информационных технологий, которое предоставляет инструменты для преобразования информации в знание. Это было определено как поиск интересных моделей и важные закономерности в больших массах информации. Говоря об эксплуатации информации, основанной на интеллектуальных системах, речь идет конкретно о применении методов интеллектуальных систем для обнаружения и перечисления закономерностей, присутствующих в информации. Методы, основанные на интеллектуальных системах, позволяют получить результаты анализа массы информации, которой обычные метод не достигают» [10].

Предлагаемая методология состоит из комбинации набора методов для процессов использования информации, основанных на эвристической интеркаляции различных алгоритмов интеллектуальных систем: алгоритмы индукции C4.5, Самоорганизующиеся Карты Кохонена (СКК) и байесовские сети доверяя (БСД), которые позволяют -соответственно- обнаружение правил поведения, обнаружение групп и обнаружение значимых атрибутов соответственно.

Апостериорные также позволяют процесс эксплуатации путем объединения алгоритма на разных этапах, например, обнаружение правил членства в группах (СКК и C4.5), взвешивание правил поведения (C4.5 и БСД), взвешивание членства в группах (СКК и БСД) и. т. д. Чтобы углубиться в тему, краткое описание каждого алгоритма приведено ниже.

Из инструмента интеллектуального анализа данных пробуются различные алгоритмические комбинации, согласно Бритос и Гарсия Мартинес [10], исследователь информации может найти следующие стратегии, основанные на его собственной эвристике расследования.

С таким образом можно получить информацию и полезные знания при меньших энергозатратах и времени созерцания со стороны аналитика.

В следующей таблице 2.1 приведено описание алгоритмов каждого метода интеллектуального анализа данных, рассматриваемого в этом анализе.

Таблица 2.1 – Методы интеллектуального анализа данных и из свойства.

Наименование метода	**Тип алгоритм**	**Ожидаемого результата**	**Тип автоматического обучения**
Алгоритм С4.5	Алгоритм индукции	Обучения деревья решения (ОДР)	Обучение с учителем
Самоорганизующаяся карта Кохонена (СКК)	Искусственная нейронная сеть (кластеризации)	Групповое обнаружение	Обучение без учителя
Байесовский сет доверия (БСД)	Скрытая Марковская модель	Графовая вероятностная модель (ГВМ)	

2.2 Принцип работы модели семантического поиска

Под эвристикой понимают совокупность приёмов и методов, облегчающих и упрощающих решение познавательных, конструктивных, практических задач [1].

Эвристический подход к поиску информации посредством процессов ЭИ основан на уникальной и оригинальной природе полезных знаний. Это так, поскольку каждая база данных и перспектива поднятой проблемы будут зависеть от ситуации и конкретного времени.

Основываясь на принципе работы ЭИ, можно сразу структурно разделить его на пять (5) типов, в пяти эвристических приближениях. Как описано в таблице 2.2 ниже, различные стратегии эксплуатации будут проверяться в соответствии с имеющимися у них данными. Ниже приведены пять подходов к ЭИ.

Таблица 2.2 – Сочетание методов интеллектуального анализа данных для эксплуатации информации.

1) Групповое обнаружение	БД → СКК → АК → БД
2) Обнаружение правил поведения	БД → С4.5 → ОДР
3) Обнаружение правил членства в группах	БД $\xrightarrow[\text{априорный}]{}$ СКК → АК БД $\xrightarrow[\text{апостериорный}]{}$ С4.5 → ОДР
4) Взвешивание правил поведения	БД $\xrightarrow[\text{априорный}]{}$ С4.5 → ОДР БД $\xrightarrow[\text{апостериорный}]{}$ БСД → ГВМ
5) Взвешивание членства в группах	БД $\xrightarrow[\text{априорный}]{}$ СКК → АК БД $\xrightarrow[\text{апостериорный}]{}$ БСД → ГВМ

Каждый из процессов использования информации может быть описан следующим образом.

2.3 Архитектура разрабатываемой модели семантического поиска

Архитектура системы поиска модели описана ниже для каждой эвристической стратегии.

Групповое обнаружение

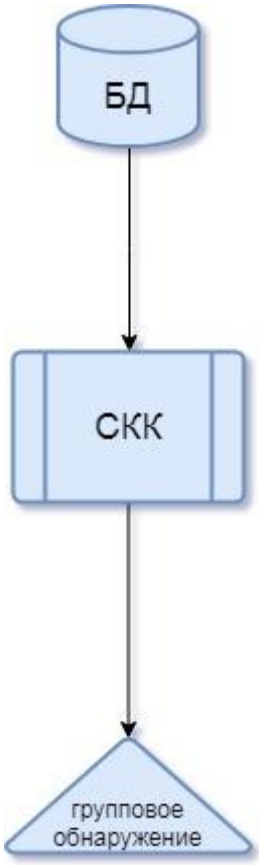

Рис. 4. Архитектура для группового обнаружения

Подход группового обнаружения сочетает в себе следующие методы добычи данных, в следующих шагах.

Групповое обнаружение	БД → СКК → АК → БД

Предыдущий Рисунок 4 архитектуры процесса может быть описан в псевдокоде следующим образом.

- Ввод данных - БД
  - Применить нейронную сеть СКК
    - Групповое обнаружение
- Выход - конец процесса

В первом аппроксимации с использованием метода кластеризации СКК подчеркивается, что обнаружение групп имеет два результата:

I. Обнаружены новые группы, которые предоставляют новые знания в опрошенной базе данных
II. Атрибут, который идентифицирует новую группу, может использоваться как атрибут класса, если он имеет значение и значение для себя.

По этой причине при поиске подходящей эвристической стратегии, которая обладает максимально возможным получением полезной информации, стратегия группового обнаружения обычно сочетается с остальными подходами.

Обнаружение правил поведения

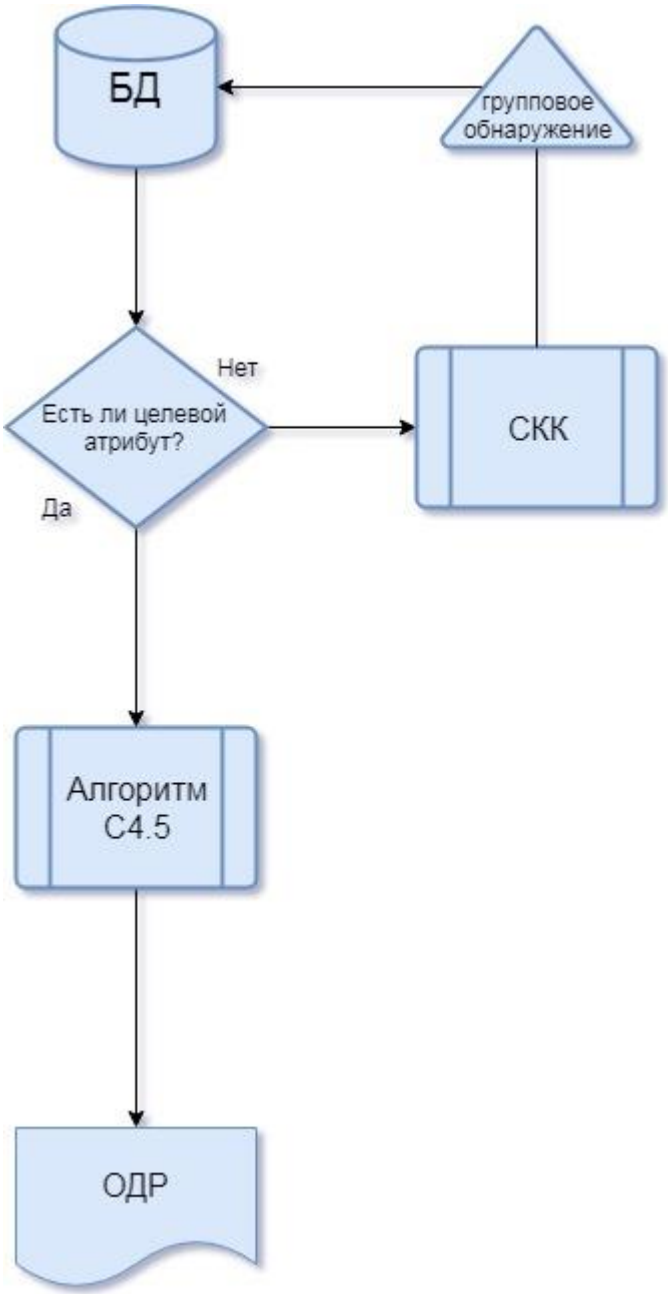

Рис. 5. Архитектура обнаружение правил поведения

Подход к обнаружению правил поведения сочетает в себе следующие методы добычи данных, в следующих шагах.

Обнаружение правил поведения	БД → C4.5 → ОДР

Предыдущий Рисунок 5 архитектуру процесса может быть описан в псевдокоде следующим образом.

- Ввод данных - БД
  - ♦ Есть ли класс или целевой атрибут?
    - Да
      - ● Применить алгоритм C4.5
        - ■ Выходные информации - ОДР
      - Выход - конец процесса
    - Нет
      - ● Применить нейронную сеть СКК
        - Δ Групповое обнаружение
- Выход - конец процесса

Обнаружение правил членства в группах

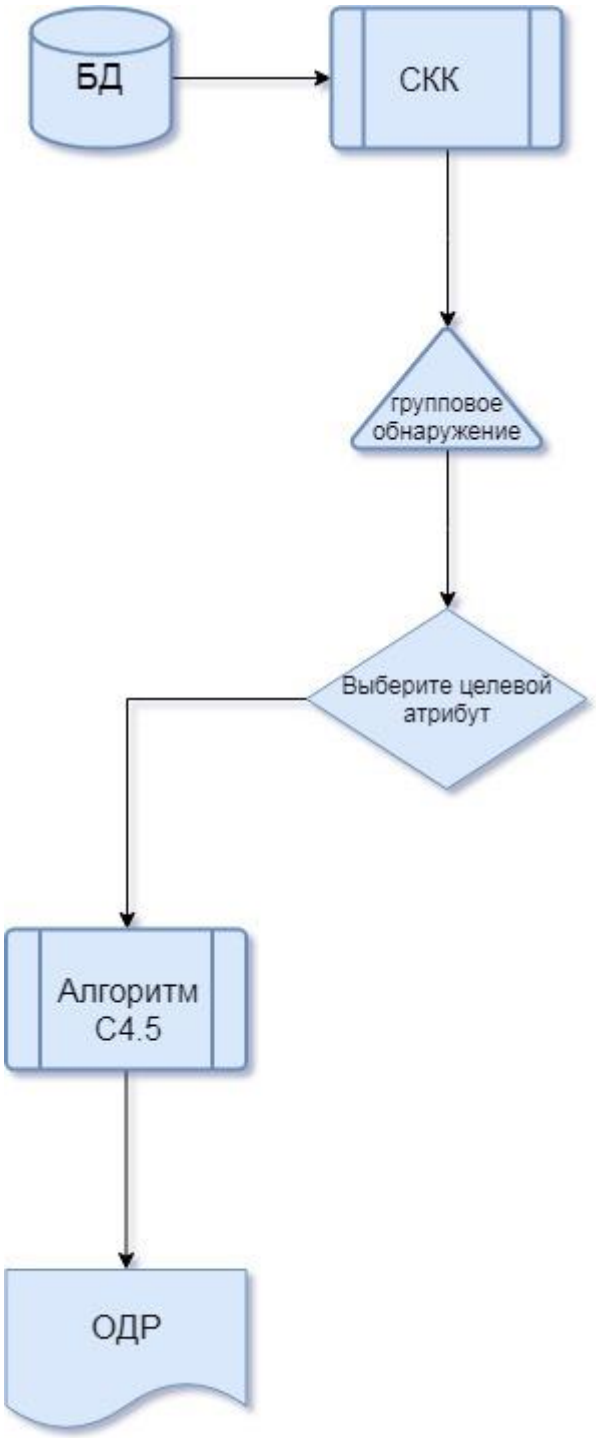

Рис. 6. Архитектура обнаружение правил членства в группах

Подход к обнаружению правил членства в группах сочетает в себе следующие методы добычи данных, в следующих шагах.

Обнаружение правил членства в группах	БД $\xrightarrow[\text{априорный}]{}$ СКК → АК БД $\xrightarrow[\text{апостериорный}]{}$ С4.5 → ОДР

Предыдущий Рисунок 6 архитектуру процесса может быть описан в псевдокоде следующим образом.

- Ввод данных–БД
  - Применить нейронную сеть СКК
    - Δ Групповое обнаружение
      - ♦ Есть ли класс или целевой атрибут?

      Да
      - Применить алгоритм С4.5
        - ■ Выходные информации – ОДР
- Выход - конец процесса

Взвешивание правил поведения

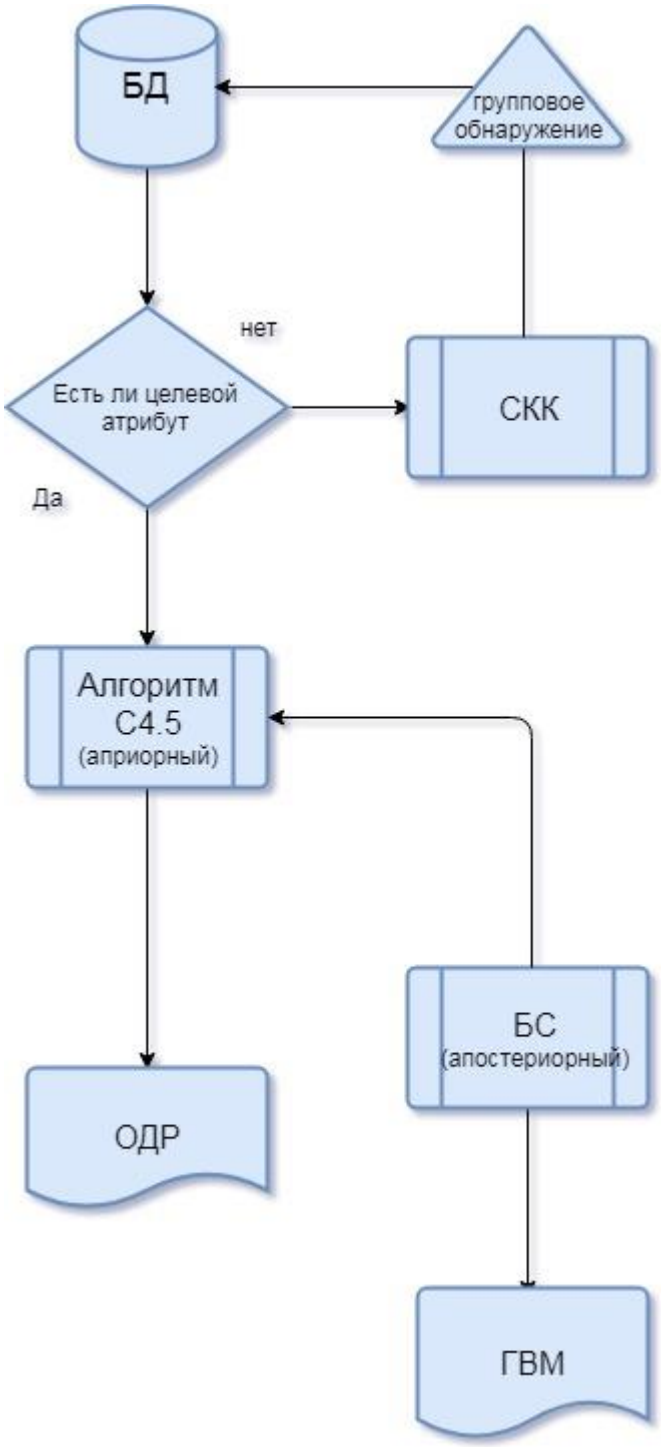

Рис. 7. Архитектура взвешивание правил поведения

Подход к взвешиванию правил поведению сочетает в себе следующие методы добычи данных, в следующих шагах.

Взвешивание правил поведения	БД $\xrightarrow[\text{априорный}]{}$ C4.5 → ОДР БД $\xrightarrow[\text{апостериорный}]{}$ БСД → ГВМ

Предыдущий Рисунок 7 архитектуру процесса может быть описан в псевдокоде следующим образом.

- Ввод данных - БД
  - ♦ Есть ли класс или целевой атрибут?

    Да
    - Применить алгоритм C4.5 (априорный)
      - ■ Выходные информации - ГВМ
    - Применить БСД (апостериорный)
      - ■ Выходные информации – ОДП

    Нет
    - Применить нейронную сеть СКК
      - Δ Групповое обнаружение
- Выход - конец процесса

Взвешивание членства в группах

Рис. 8. Архитектура взвешивание членства в группах

Подход к взвешиванию членства в группах сочетает в себе следующие методы добычи данных, в следующих шагах.

Взвешивание членства в группах	БД $\xrightarrow[\text{априорный}]{}$ СКК → АК БД $\xrightarrow[\text{апостериорный}]{}$ БСД → ГВМ

Предыдущий Рисунок 8 архитектуру процесса может быть описан в псевдокоде следующим образом.

- Ввод данных – БД
  - Применить нейронную сеть СКК (априорный)
    - Δ Групповое обнаружение
  - Применить БСД (апостериорный)
    - Выходные информации – ГВМ
- Выход - конец процесса

3. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

3.1 Выбор средств разработки

Для проведения исследования в рамках данной работы было выбрано свободно-распространяемое ПО RapidMiner для анализа данных и машинного обучения. Данное ПО реализовано на языке программирования Java - объектно-ориентированном языке программирования.

Существует множество систем ПО для интеллектуального анализа данных - как общие инструменты сбора данных, так и более частные инструменты для обработки специализированных типов данных. Ниже перечислены самые известные из них:

- Weka – ПО для анализа данных и машинного обучения.
- RapidMiner –наиболее продвинутая (вместе с Weka), open-source система, основанная на языке Java.
- KNIME – ПО для анализа данных и машинного обучения.
- Orange– ПО для анализа данных и машинного обучения.
- Tanagra – ПО для анализа данных и машинного обучения (бесплатная система).

ПО «Rapid Miner» пользовательская интерактивная среда для процессов машинного обучения и интеллектуального анализа данных. Это открытый исходный код, бесплатный проект, реализованный на Java.

Он представляет собой модульный подход к проектированию даже очень сложных задач - модульная концепция оператора, которая позволяет проектировать сложные вложенные цепочки операторов для огромного числа задач обучения. «Rapid Miner» использует XML для описания процессов обнаружения знаний в деревьях операторов.

«RapidMiner» имеет гибкие операторы для ввода и вывода данных в разных форматах файлов. Он содержит более 100 схем обучения для задач классификации, регрессии и кластеризации.

В дальнейшем будет использоваться ПО «RapidMiner» по следующим причинам:

I. Данная система является наиболее продвинутой (вместе с Weka) по сравнению с остальными.

II. Бесплатный и свободное ПО для анализа данных.

III. У автора есть большой опыт работы с данной системой.

3.2 Используемая база данных

База данных соответствует 1019 экземплярам недвижимости заявленным аргентинскими законодателями с 2003 до 2012. Источник открытой базы данных основан на 843 аффидевиты (испан. Declaraciones Juradas – заявление под присягой) или присяжное заявление государственного должностного лица от 313 законодателей депутаты и сенаторы аргентинского дума.

Хотя любая база данных любого рода могла бы использоваться, публичная база данных государственных служащих с гражданской ответственностью имеет особую полезность, потому что информация, получается от тех, кто отвечает за организацию общества, его поведение и его этическое мировоззрение.

В практическом анализе были взяты следующие девять атрибутов.

```
имущества      (ano, destino,origen, pais,
аффидавита =   titular_dominio, vinculo,
               superficie,val_decl,valor_patrim)
```

(детали каждого атрибута можно увидеть в глоссарии атрибутов приложения)

В таблице 3.1 отображается каждая переменная/атрибуты имеет следующие числовые или буквенные характеристики в своем составе.

Таблица 3.1 – Атрибут характеристики

Наименование Атрибут	**Тип входных переменных**
Ano	Количественные
Nombre	Неколичественные
Destino	Неколичественные
Origen	Неколичественные
País	Неколичественные
titular_dominio	Неколичественные
vinculo	Неколичественные
superficie	Количественные
val_decl	Неколичественные
val_patrim	Количественные

Все атрибуты связаны с характеристиками каждого имущества депутатов и сенаторов думы аргентины.

Все атрибуты связаны с характеристиками каждого имущества депутатов и сенаторов Аргентинской Думы, особенно с декларацией стоимости каждой недвижимости, ее площади, годовой декларации, происхождение средств для приобретения имущества, назначение использования, которому назначена недвижимость владения собственностью и если владелец имеет семейную связь с законодатель.

После проверки пяти стратегий использования информации и попытаться обнаружить новые группы или скрытые атрибуты класса, была выбрана комбинация методов ЭИ аналогичных «взвешиванию правил поведения».

Аналогично Рисунку 7 в следующей схеме принята эвристическая стратегия.

3.3 Принятая эвристическая стратегия

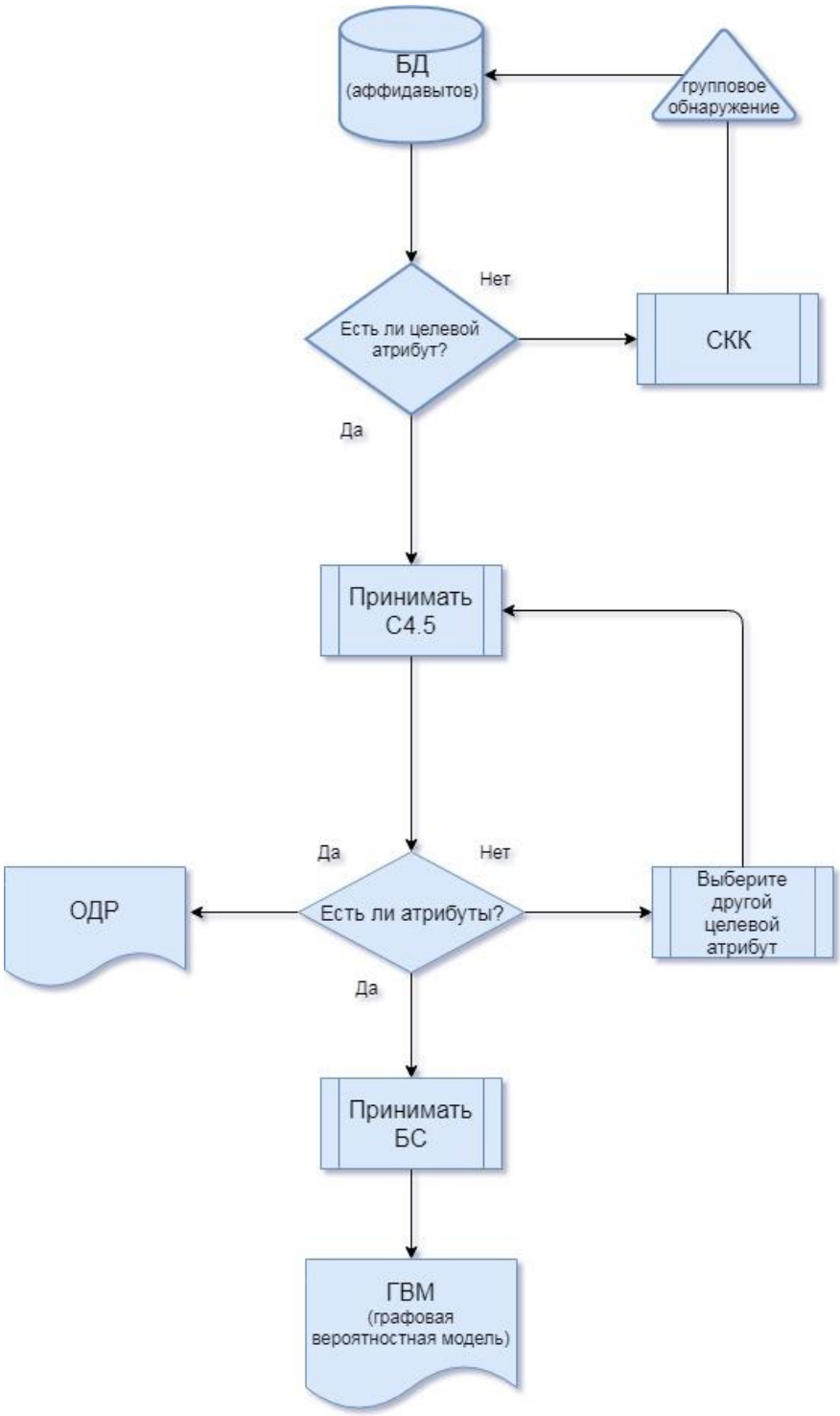

Рис. 9. Принятая эвристическая стратегия – взвешивание правил поведения

Предыдущий Рисунок 9 может быть описан в псевдокоде следующим образом.

- Ввод данных–БД
  - Определить атрибут класса
    - ♦ Есть ли класс или целевой атрибут?

    Нет
    - Применить нейронную сеть СКК
      - Δ Групповое обнаружение
        - ♦ Атрибута класс не определено выход.
      - Выход - конец процесса без получения Информации

- Ввод данных - БД
  - Определить атрибут класса
    - ♦ Есть ли класс или целевой атрибут?

    Да
    - Применить нейронную сеть СКК
      - Δ Групповое обнаружение
    - ♦ Атрибута класс определено
    - Применить алгоритм С4.5
      - ♦ Есть ли значительный прирост информации?

      Да
      - ■ Выходные информации ОДР

      Да
      - Применить БСД
        - ■ Выходные информации ГВМ

      Нет – выбрейте другой атрибут
- Выход - конец процесса

После того, как тесты были пройдены, было решено испытать метод ЭИ на реальных данных, что привело к ожидаемым и эффективным результатам.

3.4 Результаты разработки системы

Автор уже имел опыт использования этой базы данных, поскольку было легко обнаружить атрибуты объекта, которые описаны ниже в таблице 3.2. Аналогичным образом, подход «обнаружения группы» был применен в случае обнаружения кластера, который мог бы служить атрибутом класса.

В том, что не было найдено никаких новых целых атрибутов, предустановленный атрибут «`val_decl`» был взят как атрибут класса. Неколичественные атрибут, характеристики которого описаны ниже.

Таблица 3.2 – Характеристики атрибута класса (цель)

Атрибут класса (val_decl)	**Объявленная стоимость имущества относительно его рыночной и/или фискальной стоимости.**	**Обозначает/ Указывает**
Valor de Mercado	Рыночная стоимость	Нормальное поведение
Valor Fiscal	Фискальная стоимость	Подозрение в уклонении от уплаты налогов
Valor Subfiscal	Субфискальная стоимость	Подозрение в уклонении от уплаты налогов
Sin Datos	Не объявляет	Уклонение от уплаты налогов

В налоговых декларациях в Аргентине, заявление о налоговой стоимости вменяется, когда нет рыночной оценки. В Аргентине стоимость имущества может иметь два значения

I. Рыночная стоимость: реальная и представительная стоимость имущества.
II. Фискальная стоимость: базовая стоимость, назначаемая государством, как правило, намного ниже рыночной стоимости.

Следующим этапом подготовительного шага является запуск алгоритм C4.5, по которому было получено следующее дерево решений фигуры 10.

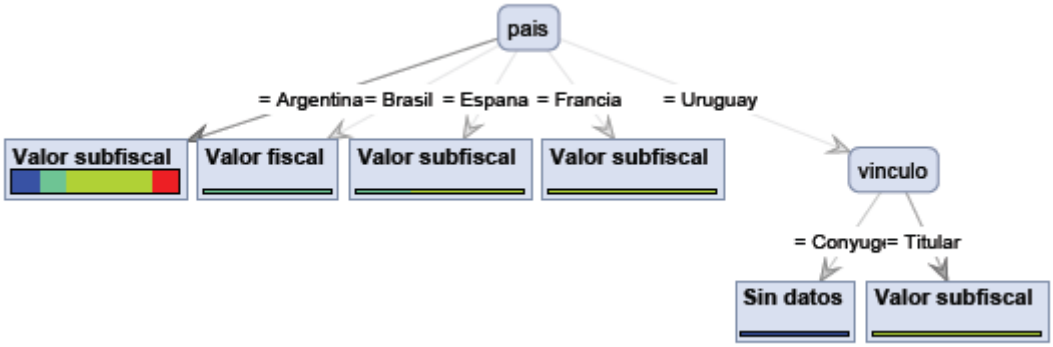

Рис. 10. Дерево решений из алгоритма индукции C4.5

Изучая результаты и полученное дерево решений, атрибуты «pais» (страна) и «vinculo» (отношения с владельцем имущества) определяются как важные для объяснения атрибута класса.

Запускаемые фигура, после выполнения разработанного шаги с атрибутом класса «объявленное значение»имущество законодателя (val_decl).

Следующим этапом подготовительного шага является запуск БСД который был сделан из следующих атрибутов.

```
имущества              (destino, origen, pais,
аффидевита =          vinculo,val_decl)
```

Учитывая, что эти атрибуты были наиболее значительными и имеют логическую, возможную и заслуживающую доверия структуру для понимания коррупционного поведения некоторых законодателей, как видно на Рисунке 11.

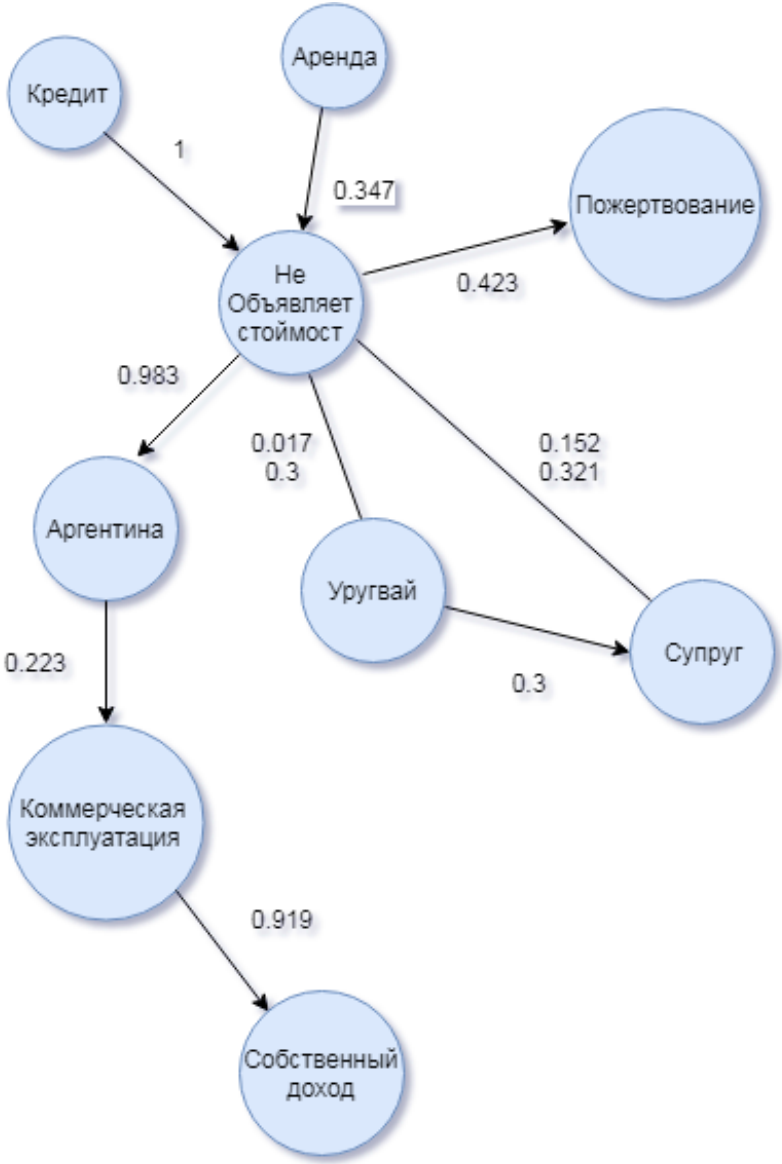

Рис. 11. Карта сети убеждений из байесовских сетей - ГВМ

Как видно из приведенных графиков (Рисунки 11 и 12) наблюдается наличие коррупционного поведения в аргентинской думе в связи с не декларированием стоимости их объектов недвижимости (в таблицах Б.1, Б.2, Б.3, Б.4 и Б.5 приложения результаты условных вероятностей могут быть пересмотрены для всех факторов).

Все законодатели, которые берут кредит, не декларируют стоимость своего имущества, и многие, кто использует свою собственность для сдачи в аренду.

Значительная часть тех законодатели, которые не декларируют стоимость своих имущества, делают пожертвования или дарение.

Те законодатели, которые владеют своей собственностью в Аргентине и не заявляют о своей стоимости, склонны отдавать эти товары в коммерческую эксплуатацию.

Те законодатели, которые владеют имуществом в Уругвае (классический курорт аргентинцев, а также налоговый рай южного конуса), зарегистрированным на имя их супруги, склонны не декларировать свою ценность.

В свою очередь, происхождение средств, которые использовались для приобретения недвижимости, предназначенной для коммерческой эксплуатации, происходит главным образом из собственного дохода законодателя, которому, по-видимому, является приоритетным направлять усилия на коммерческую деятельность вместо законодательного.

В той мере, в какой и когда был получено полезную информацию из эксперимента, подтверждается, что показания последовательны, а также показания БД (аффидевытов) согласуются и соответствуют постулатам (ООИ.1, ООИ.2, ООИ.3) следовательно $\sigma \subset$ БД(аффидевытов) т. е. семантический система в БД.

ЗАКЛЮЧЕНИЕ

В ходе данной работы был спроектирован и разработан новый метод поиска полезной информации в наборе разнородных данных. Данный метод применим для выявления полезной информации для принятия решений любого рода, равно как для открытия новых знаний в сложных системах. Путем внедрения методов эксплуатации информации было достигнуто сокращение времени и уменьшение ресурсоемкости процесса поиска семантической информации в базе данных.

В ходе работы были выполнены все поставленные задачи:

I. Произведен обзор проблем философии информации в современных условиях.
II. Проведен обзор и анализ систем эксплуатации информации для выявления возможных способов проектирования системы поиска информации.
III. Проведено комбинирование трех алгоритмов поиска и обнаружения семантических информационных систем.
IV. Произведено тестирования разработанного метода на реальном наборе данных аргентинского правительства.

Использование результатов диссертационного исследования позволяет производить поиск скрытых и неявных знаний в любой доступной базе данных.

СПИСОК ЛИТЕРАТУРЫ

[1] Википедия [Электронный ресурс]: интернет-энциклопедия – режим доступа http://ru.wikipedia.org/

[2] Демин А. В. «Математическая модель поиска объекта» — Лекция КМП № 5 — Санкт-Петербург: СПБГУ ИТМО.

[3] Иванников А. Д., Тихонов А. Н., Цветков В. Я. Некоторые аспекты теории информации — Москва: «Вестник МГОУ» Серия «Философские науки», 2012. — с. 143-145 с/. режим доступа https://vestnik-mgou.ru/Articles/Doc/3941

[4] Иванников И. В., Кулагин В. П., Мордвинов В. А., Найханова Л. В., Овезов Б. Б., Тихонов А. Н., Цветков В. Я. Получение знаний для формирования информационных образовательных ресурсов. — М: ФГУ ГНИИ ИТТ «Информика», 2008 — 440с.

[5] Кудж С. А. О философии информации. «Перспективы науки и образования», 2013. № 6. — С. 9-13с/. режим доступа https://cyberleninka.ru/article/n/o-filosofii-informatsii

[6] Кудж С. А. Коррелятивный анализ как метод познания.«Перспективы науки и образования», 2013. — №5. —С. 9-13.

[7] Лексторский В. А., Кужд С. А., Никитина Е. А.. Эпистемология, наука, жизненный мир человека. «Вестник МГТУ МИРЭА», 2014 — №2 — С 1-12. режим доступа https://rtj.mirea.ru/upload/medialibrary/650/01-lektorskii_kudj.pdf

[8] Соловьев И. В., Цветков В. Я. О содержании и взаимосвязях категорий «информация», «информационные ресурсы», «знания». «Дистанционное и виртуальное обучение», 2011. — № 6. — С. 11-21.

[9] Хлебников Г. В. Философия информации Лучано Флориди, «Метафизика» — № 4(10) — С 35-48. режим доступа http://www.intelros.ru/pdf/metafizika/2013_4/3.pdf

[10] Britos P. V., Garcia Martinez R.. «Propuesta de procesos de explotación de la información». Jujuy: XV CACIC — VI WBD. «RedUNCI», 2009. — С. 1041-1050. режим доступа http://hdl.handle.net/10915/21206

[11] Lopez-Pablos R. Elementos de ingeniería de explotación de la información aplicados a la investigación tributaria fiscal. Rosario:XLVIII Reunión Anual. «Anales de la AAEP», 2013.—С1-7. режим доступа https://aaep.org.ar/anales/works/works2013/lopez_pablos.pdf

БЛАГОДАРНОСТИ

Автор выражают искреннюю благодарность своему научному руководителю Демину Анатолию Владимировичу за интерес к данной работе и полезные советы. Также автор благодарит своих одногруппников, отличный коллектив, за помощь и бескорыстную поддержку.

ПРИЛОЖЕНИЕ А. РАЗРАБОТАННЫЙ МЕТОД ПОЙСКА – ПО RapidMiner

Указаны методы и алгоритмы, применяемые в «Rapid Miner». Сначала были найдены дополнительные группы, используя алгоритм СКК, что было безрезультативно, так как эти группы уже были найдены ранее, при помощи поиска по атрибутам класса.

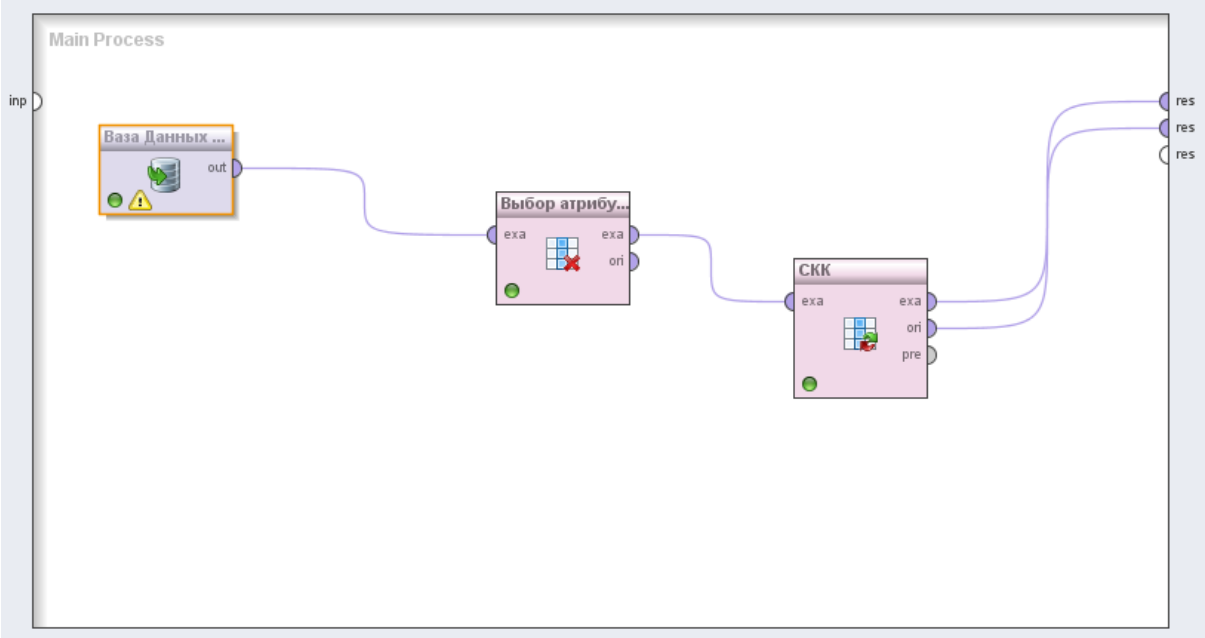

Рис. 12. Применение в «Rapid Miner» группового обнаружение.

Групповое обнаружение	БД → СКК → АК → БД

Следуя принятому эвристическому подходу (Рисунки 7, 9) была принята стратегия взвешивания правил поведения, поскольку из всех других доступных комбинаций она была получена из более полезной информации.

На практике такой подход отражен на Рисунке 13, который представляет его применение в «Rapid Miner».

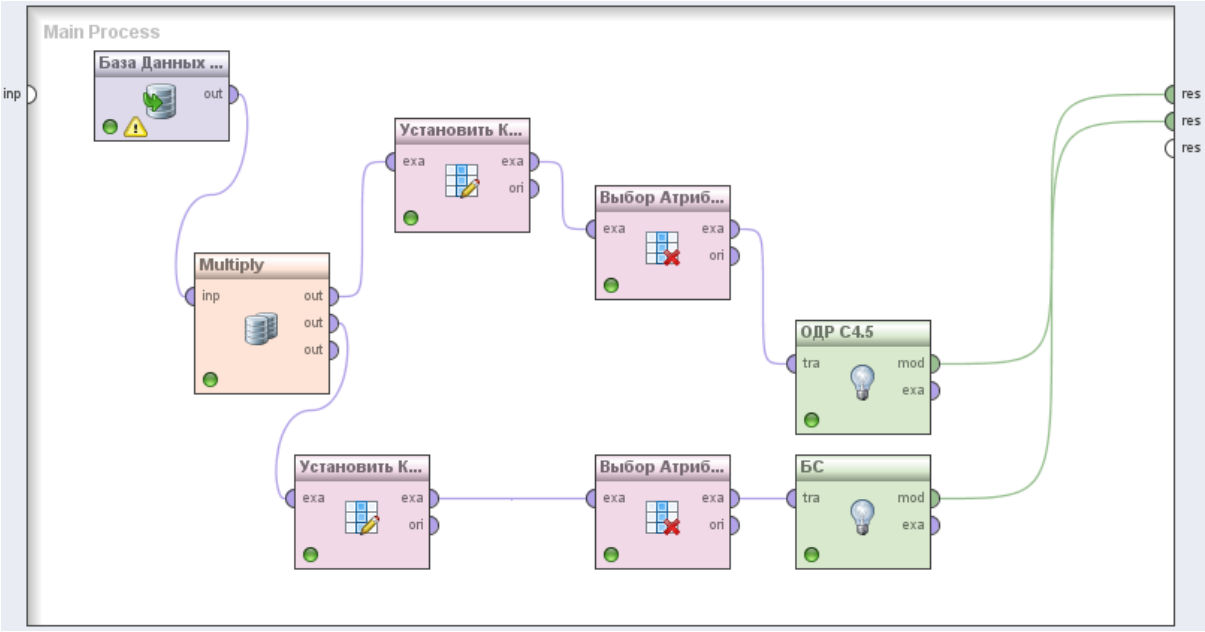

Рис. 13. Применение в «Rapid Miner» для взвешивания правил поведения.

Взвешивание правил поведения	БД $\xrightarrow[\text{априорный}]{}$ C4.5 → ОДР БД $\xrightarrow[\text{апостериорный}]{}$ БСД → ГВМ

ПРИЛОЖЕНИЕ Б. УСЛОВНЫЕ ВЕРОЯТНОСТИ БАЙЕСА ПО АТРИБУТАМ

Таблица Б.1 – Условная вероятностьповладелец недвижимости «vinculo».

атрибут	параметр	законодатель	супруг	сожитель
destino	Vivienda	0,230	0,286	0,000
destino	Alquiler	0,088	0,202	0,400
destino	Otros	0,314	0,369	0,599
destino	Nodeclara/incompleto	0,138	0,024	0,000
destino	Explotacioncomercial	0,228	0,119	0,000
origen	Credito	0,038	0,024	0,000
origen	Herencia	0,096	0,214	0,000
origen	IngresosPropios y Otros	0,015	0,036	0,000
origen	IngresosPropios	0,766	0,643	0,998
origen	IngresosConyuge	0,003	0,048	0,000
origen	Nodeclara	0,005	0,024	0,000
origen	Propiedadesanteriores	0,004	0,000	0,000
origen	Donacion	0,070	0,000	0,000
origen	Legado	0,000	0,012	0,000
origen	Usufructo	0,003	0,000	0,000
pais	Argentina	0,976	0,964	0,999
pais	Brasil	0,003	0,000	0,000
pais	Uruguay	0,008	0,036	0,000
pais	Francia	0,010	0,000	0,000
pais	Espana	0,003	0,000	0,000
val_decl	Sindatos	0,162	0,321	0,000
val_decl	Valorfiscal	0,160	0,107	0,000
val_decl	Valorsubfiscal	0,519	0,452	0,999
val_decl	Valormercado	0,158	0,119	0,000

Таблица Б.2 – Условная вероятность в стране «pais».

атрибут	параметр	Аргентина	Бразилия	Уругвай	Франция	Испания
destino	Vivienda	0,224	0,000	0,600	0,999	0,000
destino	Alquiler	0,102	0,000	0,000	0,000	0,000
destino	Otros	0,318	0,998	0,400	0,000	0,998
destino	Nodeclara/incompleto	0,131	0,000	0,000	0,000	0,000
destino	Explotacioncomercial	0,223	0,000	0,000	0,000	0,000
Origen	Credito	0,035	0,000	0,000	0,000	0,665
Origen	Herencia	0,108	0,000	0,000	0,000	0,000
Origen	IngresosPropios y Otros	0,016	0,000	0,000	0,000	0,332
Origen	IngresosPropios	0,754	0,997	0,999	0,999	0,000
Origen	IngresosConyuge	0,007	0,000	0,000	0,000	0,000
Origen	Nodeclara	0,007	0,000	0,000	0,000	0,000
Origen	Propiedadesanteriores	0,004	0,000	0,000	0,000	0,000
Origen	Donacion	0,065	0,000	0,000	0,000	0,000
Origen	Legado	0,001	0,000	0,000	0,000	0,000
origen	Usufructo	0,003	0,000	0,000	0,000	0,000
vinculo	Titular	0,913	0,999	0,700	1,000	0,999
vinculo	Conyuge	0,081	0,000	0,300	0,000	0,000
vinculo	Conviviente	0,005	0,000	0,000	0,000	0,000
val_decl	Sindatos	0,176	0,000	0,300	0,000	0,000
val_decl	Valorfiscal	0,155	0,999	0,000	0,000	0,333
val_decl	Valorsubfiscal	0,511	0,000	0,700	1,000	0,666
val_decl	Valormercado	0,158	0,000	0,000	0,000	0,000

Таблица Б.3 – Условная вероятность в относительной стоимости имущества «val_decl».

атрибут	параметр	не объявляет	фискальная стоимость	субфискальная стоимость	риночная стоимость
destino	Vivienda	0,264	0,177	0,266	0,146
destino	Alquiler	0,197	0,057	0,084	0,083
destino	Otros	0,180	0,386	0,350	0,312
destino	Nodeclara/ incompleto	0,213	0,120	0,034	0,350
destino	Explotacioncomercial	0,135	0,259	0,266	0,108
origen	Credito	0,079	0,070	0,006	0,057
origen	Herencia	0,157	0,171	0,065	0,115
origen	IngresosPropios y Otros	0,006	0,013	0,021	0,019
origen	IngresosPropios	0,595	0,702	0,888	0,554
origen	IngresosConyuge	0,000	0,019	0,006	0,006
origen	Nodeclara	0,011	0,013	0,006	0,000
origen	Propiedadesanteriores	0,000	0,000	0,008	0,000
origen	Donacion	0,135	0,006	0,002	0,248
origen	Legado	0,000	0,006	0,000	0,000
origen	Usufructo	0,017	0,000	0,000	0,000
Pais	Argentina	0,983	0,975	0,966	1,000
pais	Brasil	0,000	0,019	0,000	0,000
pais	Uruguay	0,017	0,000	0,013	0,000
pais	Francia	0,000	0,000	0,017	0,000
pais	Espana	0,000	0,006	0,004	0,000
vinculo	Titular	0,848	0,943	0,918	0,936
vinculo	Conyuge	0,152	0,057	0,072	0,064
vinculo	Conviviente	0,000	0,000	0,010	0,000

Таблица Б.4 – Условная вероятность по назначению имущества«destino».

атрибут	параметр	жилье	аренда	другие	Не объявляет	Коммерчечкая эклсплуатация
origen	Credito	0,080	0,010	0,015	0,077	0,009
origen	Herencia	0,097	0,277	0,110	0,069	0,050
origen	IngresosPropios y Otros	0,034	0,040	0,012	0,000	0,005
origen	IngresosPropios	0,748	0,614	0,834	0,415	0,923
origen	IngresosConyuge	0,013	0,020	0,006	0,000	0,000
origen	Nodeclara	0,004	0,010	0,006	0,015	0,005
origen	Propiedadesanteriores	0,017	0,000	0,000	0,000	0,000
origen	Donacion	0,008	0,000	0,012	0,423	0,009
origen	Legado	0,000	0,000	0,003	0,000	0,000
origen	Usufructo	0,000	0,030	0,000	0,000	0,000
pais	Argentina	0,937	1,000	0,969	1,000	1,000
pais	Brasil	0,000	0,000	0,009	0,000	0,000
pais	Uruguay	0,025	0,000	0,012	0,000	0,000
pais	Francia	0,038	0,000	0,000	0,000	0,000
pais	Espana	0,000	0,000	0,009	0,000	0,000
vinculo	Titular	0,899	0,812	0,896	0,985	0,955
vinculo	Conyuge	0,101	0,168	0,095	0,015	0,045
vinculo	Conviviente	0,000	0,020	0,009	0,000	0,000
val_decl	Sindatos	0,197	0,347	0,098	0,292	0,108
val_decl	Valorfiscal	0,118	0,089	0,187	0,146	0,185
val_decl	Valorsubfiscal	0,588	0,436	0,564	0,138	0,631
val_decl	Valormercado	0,097	0,129	0,150	0,423	0,077

Таблица Б.5 – Условная вероятность по происхождения имущества «origen».

Атрибут	Параметр	кредит	Наследие (herencia)	собственный доход и другие	собственный доход	доход супруга	не объявляет	недвижимость прошлого	дарение/ пожертвование	наследие (legado)	узуфрукт
destino	Vivienda	0,513	0,215	0,470	0,231	0,428	0,143	0,999	0,031	0,001	0,000
destino	Alquiler	0,027	0,262	0,235	0,080	0,286	0,143	0,000	0,000	0,001	0,998
destino	Otros	0,135	0,336	0,235	0,353	0,286	0,286	0,000	0,062	0,994	0,000
destino	Nodeclara/incompleto	0,270	0,084	0,000	0,070	0,000	0,286	0,000	0,846	0,001	0,000
destino	Explotacioncomercial	0,054	0,103	0,059	0,266	0,000	0,143	0,000	0,031	0,001	0,000
pais	Argentina	0,946	1,000	0,941	0,971	0,999	0,999	0,999	1,000	0,995	0,998
pais	Brasil	0,000	0,000	0,000	0,004	0,000	0,000	0,000	0,000	0,001	0,000
pais	Uruguay	0,000	0,000	0,000	0,013	0,000	0,000	0,000	0,000	0,001	0,000
pais	Francia	0,000	0,000	0,000	0,012	0,000	0,000	0,000	0,000	0,001	0,000
pais	Espana	0,054	0,000	0,059	0,000	0,000	0,000	0,000	0,000	0,001	0,000
vinculo	Titular	0,946	0,832	0,823	0,923	0,428	0,714	0,999	1,000	0,001	0,999
vinculo	Conyuge	0,054	0,168	0,176	0,070	0,571	0,286	0,000	0,000	0,997	0,000
vinculo	Conviviente	0,000	0,000	0,000	0,006	0,000	0,000	0,000	0,000	0,001	0,000
val_decl	Sindatos	0,378	0,262	0,059	0,137	0,000	0,286	0,000	0,369	0,001	0,999
val_decl	Valorfiscal	0,297	0,252	0,118	0,144	0,428	0,286	0,000	0,015	0,996	0,000
val_decl	Valorsubfiscal	0,081	0,318	0,647	0,606	0,428	0,428	0,999	0,015	0,001	0,000
val_decl	Valormercado	0,243	0,168	0,176	0,113	0,143	0,000	0,000	0,600	0,001	0,000

ПРИЛОЖЕНИЕ В. ГЛОССАРИЙ АТРИБУТОВ

ano:	Годовой период, к которому относится показание должностного лица.
destino:	Предназначение декларируемой собственности
origen:	Происхождение ресурсов, которые оправдывают владение или распоряжение имуществом.
país:	Страна происхождения имущества, декларируемая законодательным должностным лицом.
titular_dominio:	Владелец недвижимости, заявленный законодателем.
vinculo:	Тип отношений с владельцем и / или совладельцем объявленной недвижимости.
superficie:	Площадь в квадратных метрах заявленной недвижимости.
valor_patrim:	Объявленная стоимость недвижимости в аргентинских песо.
val_decl:	Объявленная стоимость имущества относительно его рыночной и/или фискальной стоимости.