**Концепция разработки и функционирования информационно-вычислительного портала ArcNet**

Общая задача исследования заключается в разработке концепции, включающей методы, алгоритмы и компьютерные технологии для реализации и внедрения на платформе современных программных решений целого ряда интегрированных приложений с единым интерфейсом – Арктической интеллектуальной информационной системы (АИИС) со специальным Арктическим электронным порталом ArcNet, обеспечивающего гарантированный уровень безопасности, коллективную работу и взаимодействие с офисными программами.

Основная идеология разрабатываемой системы заключается в создании обобщенной информационной среды Арктического пространства. Эта среда позволит агрегировать экономические и геополитические данные по всем Арктическим регионам, а также организовывать моментальный доступ к ним и их обработку для последующей аналитики.

Площадкой, позволяющей работать с этой системой должен выступать web-портал ArcNet. Задача этого портала – организовывать доступ из любой точки мира к необходимым данным различному кругу лиц – от частных лиц до крупных государственных организаций.

Web-портал должен иметь связь со статистической базой, а так же с двумя программными модулями, которые позволят осуществлять непосредственный поиск и агрегацию данных и их последующую обработку в соответствии с рядом заложенных функциональных методов и методик моделирования.

Укрупненный алгоритм разработки данной системы, предлагаемый автором, делится на несколько основных этапов:

1. Разработка программного продукта 1 (ПП1), направленного на поиск и систематизацию данных
2. Разработка web-портала, организовывающего доступ к данным
3. Разработка программного продукта 2 (ПП2), направленного на обработку данных и формирование прогностических моделей по заданным исследователем требованиям
4. Организация взаимосвязи между всеми элементами системы

Основные ограничения, возникающие при реализации данного алгоритма:

1. Необходимость классификации данных и выявления ограничения их типа. Совершенно очевидно, что для адекватной работы проектируемой системы данные должны быть представлены в четком количественном виде. Однако, предполагается и возможное наличие неколичественных данных, которые впоследствии должны быть обработаны ПП1 и представлены в унифицированном виде.
2. Систематизируемые данные в должны быть помещены в базы данных, в дальнейшем подключаемые и web-порталом и ПП2, направленным на обработку этих данных. Соответственно базы данных должны быть либо иерархические, либо реляционные. Второй вариант с точки зрения развития web-портала более перспективно. Так же допускается создание сложных комбинированных баз данных.
3. Предполагаемое большое количество данных необходимо грамотно систематизировать, а также выстраивать ключевые связи между ними для упрощения дальнейшего поиска и обработки. В связи с этим предполагается деление данных на определенные слои. Это могут быть слои, сформированные на основе, например, географической принадлежности или же отраслевой. Возможны и иные варианты.
4. При разработке программного комплекса обработки данных встает вопрос о наборе инструментария, а именно о доступных для вычисления моделях.
5. Планируемые огромные массивы данных и сложные вычислительные программные комплексы должны быть подкреплены сопоставимыми техническими мощностями серверов.

Одной из первоочередных задач является разработка таксономии специализированного Арктического электронного портала. Общий структурный вид АИИС, разработанный автором, имеет вид, представленный на рис. 1



*Рис. 1. Общая структурная схема АИИС*

Из рисунка видно, что разрабатываемая система должна быть доступна из сети интернет. Для полноценной работой с системой необходимо обладать логином и паролем. Пользователи, зарегистрированные в системе, могут быть разделены на группы, с разными правами доступа и возможностями в зависимости от типа учетных записей: частные лица, организации, государственные структуры. Так же, предполагается наличие администраторов и модераторов – участников системы с возможностями ее настройки и корректировки. Возможна организация гостевого доступа с минимальным набором ознакомительных возможностей.

Уровни доступа предполагают разграничение пользовательских возможностей по глубине искомой информации, по ее географическому представлению или представлению в виде слоев. Частное лицо может иметь доступ к минимуму функций и отображению основной экономической информации по субъектам. Организации могут иметь возможность находить расширенную информацию по конкретным секторам (отраслям) экономической деятельности, а также производить некоторые вычисления и прогнозирование на основе этих данных. Уровень муниципалитетов предполагает доступ ко всей статистической информации в пределах этого муниципалитета по всем секторам экономической деятельности. Доступны также и средства моделирования и прогнозирования.

Рассмотрим работу всей системы более детально.



*Рис. 2. Принцип работы ПП1. Блоки I, II. Поиск, агрегация и представление данных*

Для этого отразим принципы функционирования каждого ее блока в отдельности. Исходя их рисунка 2 можно разобрать предназначение и принцип работы ПП1.

Ключевая особенность этого программного продукта заключается в том, что это самостоятельное программное решение, позволяющее в автоматизированном порядке собирать статистические данные по указанным запросам из различных источников сети интернет и в дальнейшем ее группировать и систематизировать. Для выполнения этой функции автор предлагает использовать математические методы, основанные на принципах обучаемой нейронной сети. Этот подход с одной стороны позволит упростить поисковый алгоритм, а с другой стороны даст широкий спектр возможностей для дальнейшей корректировки и точечной настройки обработки данных. Помимо этого, процесс обучения на первых порах будет способствовать максимально точному воспроизведению искомой информации.

Фактически, в ПП1 администратором системы задаются ключевые запросы для поиска требуемой информации. ПП1 обрабатывает запросы и в соответствии с указанными ключевыми словами, метками или областями осуществляет поиск необходимой информации в сети интернет.

Найденная информация загружается из ресурсов сети интернет и помещается в предварительные базы данных для дальнейшего анализа и структуризации.

В качестве структуры базы данных автор полагает использование баз реляционного типа. Это обусловлено тем, что для реляционной базы данных под записью понимается строка прямоугольной таблицы. Элементы записи образуют столбцы этой таблицы (поля). Все элементы в столбце имеют одинаковый тип (числовой, символьный и др.), а каждый столбец – неповторяющееся имя. Одинаковые строки в таблице отсутствуют. Основное преимущество таких баз данных – наглядность и понятность организации данных, скорость поиска нужной информации, что является немаловажным аспектом для конечно АИИС.

ПП1 автоматически согласно алгоритма разделяет загруженные данные на три типа:

* Структурированные
* Слабо структурированные
* Не структурированные

и помещает их в разные таблицы базы данных. Обязательным является сохранение источника копированных данных. Это позволит избежать дальнейшего уточнения полученных данных, путем объединения из разных источников.

Важно отметить, что если заданный запрос совпадал с ранее обработанным, то данные по этому запросу уже есть в базе данных. ПП1 после обработки вторичного запроса должен сравнить данные с уже имеющимися и консолидировать их.

Администратор может проверять полученную структуру данных.

Структурированные данные помещаются в конечную базу данных

ПП1 анализирует данные по степени их надежности, путем обработки источников копирования. Могут быть определены три типа степени надежности:

* Достоверные данные: числовые и нечисловые данные из проверенных источников (например, ресурсы гос. статистики)
* Слабо достоверные: числовые данные из непроверенных источников
* Оценочные: нечисловые данные, базирующиеся на мнениях, оценках и т.д.

В связи с указанной шкалой ПП1 после анализа сформированной БД каждому блоку данных присваивает ярлык, отражающий степень достоверности. Администратор может отслеживать присваиваемые ярлыки и по необходимости корректировать их

Полученная БД коммутируется к Web-интерфейсу портала ArcNET

Пользователь может просматривать информацию из БД путем введения требуемых запросов или выбора из предложенных вариантов согласно уровням доступа учетной записи.

Как было отмечено ранее, базы данных рекомендуется использовать реляционного типа. Помимо этого, было бы целесообразно разделять эти базы данных на отдельные слои, обуславливаемые регионами Арктической зоны. Если говорить более расширенно о построении слоев, то автор предлагает условно разделить слои на три уровня, представленные на рис. 3.



*Рис. 3. Принцип разделения данных на слои*

Первый и самый низший слой – это уровень отдельно взятой области в пределах сектора экономической деятельности. Второй – сами сектора. Третий слой при таком подходе будет отражать уровень муниципалитетов, а четвертый, соответственно, уровень регионов. В итоге все сводится к последнему и самому крупному слою – уровню Арктической зоны в целостности. В дальнейшем допускается добавление слоев, сформированных для уровней приарктических регионов, иных регионов Российской Федерации, возможно, для уровней других стран, имеющих непосредственное отношение к развитию Арктического пространства.

Помимо программного продукта, направленного на поиск и агрегацию данных система должна включать в себя программный продукт, позволяющий проводить различные вычисления над собранными данными.



*Рис. 4. Принцип работы ПП2. Блок III. Обработка данных и проведение вычислений*

ПП2, представленный на рисунке 4 обладает некоторым вычислительным функционалом, обусловленным подключаемыми методами и моделями статистической обработки.

Пользователь, с учетом своего уровня доступа, через электронный web-портал ArcNet путем формирования запроса ставит задачу на обработку данных и вычисление характеристик на основе предлагаемых вариантов методик и моделей. Web-интерфейс обрабатывает запрос и посылает его на выполнение программному продукту 2 (ПП2).

В соответствии с указанными пользователем требованиями ПП2 обращается к базе данных с целью получения необходимых исходных данных. Данные извлекаются из соответствующей базы данных и попадают в блок расчета ПП2.

В соответствии с указанными пользователем требованиями ПП2 обращается к аналитическому блоку с целью подключения необходимых моделей для дальнейших вычислений. В блок расчета подключаются требуемые модели, после чего производится расчет и итог записывается в таблицу результатов.

Результат загружается в динамическую страницу Web-интерфейса и отображается на обновляемой странице пользователю. В дальнейшем пользователь имеет возможность напечатать или скопировать полученный результат. Помимо этого, web-интерфейс обязан иметь модуль экспорта полученных результатов в формат традиционных офисных приложений, таких как MS Word или MS Excel.

Администратор на уровне доступа непосредственно к ПП2 имеет возможность корректировать и дополнять методиками и моделями аналитический блок. Так же он может настраивать ПП2. Для качественной и быстрой работы ПП2 особенно важно иметь хороший запас ресурса по технологической мощности. Весь описанный выше процесс от момента создания запроса до получения результата не должен составлять более нескольких секунд.

Описывая возможности пользователей системы стоит обратить особенное внимание на клиентов системы и администраторов. К клиентам системы можно отнести любого пользователя вне зависимости от уровня доступа. Клиенты имеют доступ только к web-интерфейсу и могут осуществлять только доступные им, заранее созданные администраторами функции. Как видно из рисунка 5 файла «Дополнительные материалы к отчету.docx» пользователь имеет ограниченный полномочия по допуску к ресурсам разрабатываемой системы. Фактически, при наличии интернета пользователь может иметь доступ к web-интерфейсу базы данных, а также проводить вычисления путем выбора методов расчета из предлагаемых в web-клиенте. Web-клиент самостоятельно связывается с требуемыми базами данных и вычислительным комплексом ПП2. Все вычислительные операции производятся по заранее заданным алгоритмам без участия пользователя. В конечном итоге, после введения запросов/запуска вычислений пользователь в web-интерфейсе увидит результаты запросов/вычислений.

В рисунке 6 файла «Дополнительные материалы к отчету.docx» отражены возможные воздействия администратора на различные узлы системы. Очевидно, что администратор способен управлять как самим web-интерфейсом, и проводить все манипуляции, доступные пользователю, так и имеет доступ к программным продуктам ПП1 и ПП2 напрямую. Администратор имеет доступ к web-оболочке не только через интернет, но и локально.

Разработка АИИС требует согласованных действий в различных областях – начиная от алгоритмизации и заканчивая программированием и тестированием системы.