

**Министерство образования и науки Российской Федерации
МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ
(государственный университет)
ФАКУЛЬТЕТ УПРАВЛЕНИЯ И ПРИКЛАДНОЙ МАТЕМАТИКИ
КАФЕДРА ИНФОРМАТИКИ
(Специализация «Современные компьютерные технологии»)**

Когнитивный процессинг социальных сетей

**Бакалаврская диссертация
или: Выпускная квалификационная работа
студента 673 группы
Парамонова Сергея Валерьевича**

**Научный руководитель
Рыков В.В. к.фил.н**

г. Долгопрудный

2010

Table of Contents

Глава 1.	3
§1.1 Постановка задачи. Предпосылки.....	3
§1.2 Постановка задачи	4
Глава 2.....	5
§2.1.	5
§2.2 Существующие решения.	6
§2.3 Японский подход к решению проблемы.	8
§2.4 Почему применение семиотики действительно критично в данном контексте.	9
Глава 3	12
§3.1 JDBC-ODBC аналогия.	12
§3.2 Semantic Web.	14
Глава 4. Модель.....	17
§4.1 Схема.....	17
§4.2 Элементы модели.....	18
Сетей.....	18
Операции над сетями:.....	18
Метаструктура сети.....	20
Глава 5. Алгоритм и метод.....	22
§5.1 Составляющие алгоритма.....	24
Политика присутствия.....	24
Создание метаструктуры.....	24
Интеграция на основе открытой платформы.....	24
Глава 6.....	26
§6.1 Программа.....	26
Глава 7. Эксперимент.....	27
§7.1 Численный эксперимент.....	27
§7.2 Результат.....	28
§7.3 Дальнейшее направление исследований	28
Список литературы.	30

Глава 1.

§1.1 Постановка задачи. Предпосылки

В эру информационных технологий довольно типичным является работа над проектом, в котором принимает участие интернациональная команда специалистов из различных областей[1] .

Именно в такие периоды наиболее остро встает проблема взаимной работы разнородного коллектива. Современные бизнес-процессы требуют минимизации количества ошибок, связанных с выполнением задач; минимизации скорости выполнения и общей стоимости реализации и поддержки проекта. В подобных случаях мы имеем дело, с так называемым Knowledge Gap (следуя введенной компанией IBM терминологии). Knowledge Gap - это проблема нехватки единой системы символов, необходимой для совместной работы команды (collaborative work). Иначе говоря, специалисты различных уровней исполнения и различных предметных областей, при выполнении общей части проекта имеют некоторый коэффициент снижения уровня исполнения за счет недопонимания, вызванного отсутствием единой и общей системы коммуникации и терминологии[2].

Формальная постановка проблемы:

- 1) Рост компаний\проектов в перспективе приводит к работе в неоднородном коллективе специалистов, работающих на различных уровнях исполнения и ответственности. И проблема связывания модулей системы начинает мажорировать локальные решения по модулям и техническим трудностям инженерного характера.
- 2) Объемы информации, с которой компании\обществу и\или отдельному человеку\сотруднику приходится сталкиваться в единицу времени возрастает быстрее, чем современные системы способны адекватно их обрабатывать.
- 3) Классические методы обработки информации, сталкиваясь с задачами, преподносимыми сверхбыстроразвивающимся информационным пространством, дают результаты плохо коррелирующие с ожиданиями, растущими в соответствии с ростом объемов информации.
- 4) Область применения методов, в большинстве своем, а точнее их результаты могут быть корректно интерпретированы лишь узким спектром специалистов.

Наука занимающаяся разработкой решения для подобного круга задач именуется методологией. Данная наука, фактически опираясь на наивную предметную, область переносит некоторую модель(во многом схожую с привычной математической) на некоторую конкретную\реальную область.

Можно утверждать, что на основании найденных взаимосвязей (чаще всего они носят характер закономерностей) пытаемся выработать иерархию отношений между объектами. А

в последствии в случае успешной аппроксимации исходной задачи разработать общий шаблон решения для данного класса задач. Стоит особо подчеркнуть, что данные и сама модель не являются статичными, то есть возможно расширение предметной области, а как следствие модификация или коренной пересмотр, полученного шаблона.

В случае с общей задачей первого типа, так называемой проблемой collaborative work, требуется поставить задачу перед специалистами различного круга деятельности на языке, на котором они смогут эффективно взаимодействовать, работая и поддерживая систему. Как типичный пример, можно указать самолет Boeing 777, в связи с тем, что объемы документации сравнимы с реальным размером самолета, в разработке и поддержании участвовали сотни людей, для поддержания требуются многомиллионные вложения. Причем темп роста стоимости и человеко-временные затраты возрастают непропорционально(нелинейно) быстро[3].

В случае задачи второго типа, можно говорить о проблеме аналитики и обработки символьных данных, связанных с ростом сети Интернет. Уже классическими примерами будут социальные сети[4].

Стоит также отметить, что задачи третьего и четвертого типа могут быть вызваны такой на первый взгляд неочевидной вещью, как корпоративная политика или кадровые тренинги, разработка единого подхода к решению проблемы, как правило обладающих свойством локализуемости(задачи или ее часть могут быть решены локально) [5].

В подобном случае, стоит указать на единство этих задач. В каждой из них необходимым элементом является наличие единой системы знаков. Поставленные задачи не новы, но до сих пор пытались использовать только внутрисистемный спектр знаков, обладающих рядом существенных недостатков, не позволяющих подойти к решению достаточно эффективно. Поэтому в данной статье будут описаны внесистемные подходы к описанию подобной ЗС.

§1.2 Постановка задачи

Распространение социальных сетей в последние 3-5 лет выявило как ряд дополнительных преимуществ для компаний в организации бизнеса, так и ряд проблемных моментов и рисков[6]. К последним можно отнести:

- слом организационной иерархии, служившей компаниям десятки лет,
- потерю производительности по причине длительного пребывания сотрудников в социальных сетях в рабочее время,
- ущерб для репутации компании, когда сотрудники обсуждают внутренние проблемы на общественных сайтах или, хуже того, конфиденциальную информацию, связанную с

клиентами, - утрату контроля за прохождением информации вне установленных каналов коммуникации,

- возможность для конкурирующих компаний сбора информации через социальные сети и позиционирования своей стратегии.

Основной является представление предложений по нахождению компромисса между рисками и преимуществами. Предложения могут быть сделаны как в плане ужесточения контроля над рисками, так и лучшего использования возможностей, предоставляемых социальными сетями. Особенный интерес представляла бы разработка руководства для компаний по установлению такого контроля над рисками, который не снижал бы креативных и инновационных возможностей использования сетей.

Формальные критерии оценки с точки зрения бизнеса:

- Каким образом предложения могут устроить консервативно настроенный бизнес?
- Насколько инновационно предложение в организационном и технологическом плане?
- Насколько практически осуществимы предлагаемые идеи?
- Насколько хорошо будут работать предложения в различных отраслях промышленности и разных национальных бизнес - культурах?
- Насколько сбалансированы предложения между усилением контроля и извлечением пользы от вхождения в социальные сети?

Данная формулировка задачи является, классическим вариантом постановки высокотехнологичной задачи от бизнеса и индустрии. Для решения которой необходимо проведения серьезного исследования не только в области современных технологий, но и формально проработанного математического аппарата.[\[7\]](#)

Глава 2

§2.1.

Методология (от греч. учение о способах;) учение о системе понятий и их отношений, — система базисных принципов, методов, методик, способов и средств их реализации в организации и построении научно-практической деятельности людей[\[8\]](#)

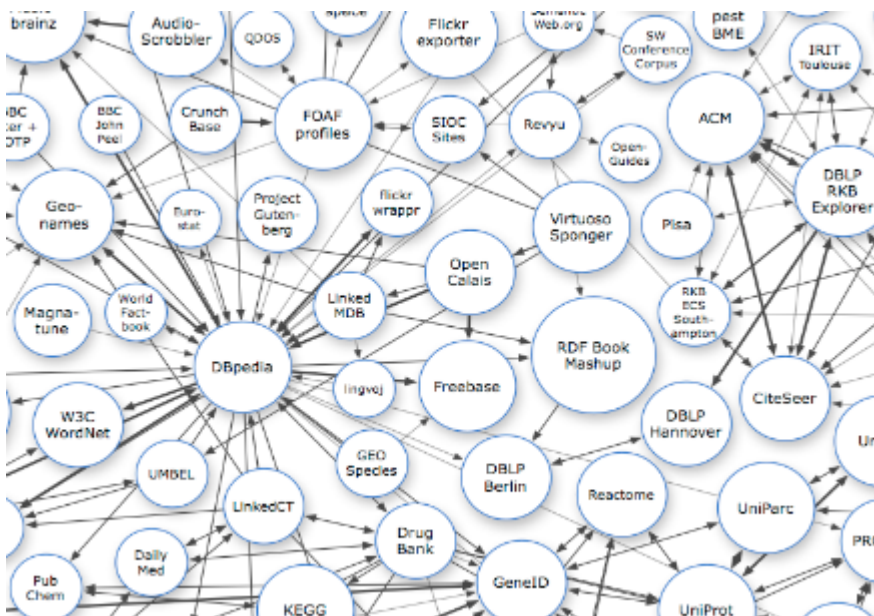
Существует несколько классических подходов к изучению задач, рассмотрим один из них.

Входные данные	Процесс обработки	Выходные данные	Описание
Не присутствуют	Не присутствуют	Не присутствуют	Данный тип задач нас не интересует, так как он не обладает никакой информацией о задаче
Не присутствуют	Не присутствуют	Присутствуют	Этот тип задач характеризуется наличие только результата,

Не присутствуют	Присутствуют	Не присутствуют	который должен быть достигнут. В англоязычной литературе именуемый Wishful Thinking
Не присутствуют	Присутствуют	Не присутствуют	В данном случае, мы имеем инструментарий и пытаемся определить класс задач, который он может решить.
Не присутствуют	Присутствуют	Присутствуют	Как и в предыдущем случае, обладаем процессом решения, но так же дополнительно имеем результат работы процесса. Поэтому это задача на проверку гипотезы.
Присутствуют	Не присутствуют	Не присутствуют	Классический тип задач, присутствует только условие задачи.
Присутствуют	Не присутствуют	Присутствуют	Задача из так называемой R&D среды, когда нужно построение процесса, строго отвечающих требования системы.
Присутствуют	Присутствуют	Не присутствуют	Задача из экспериментаторской деятельности, когда мы обладаем необходимым инструментарием для проведения эксперимента и возможностью его проведения.
Присутствуют	Присутствуют	Присутствуют	Решенная задача.
В нашем случае мы имеем дело с задачей типа R&D, когда определены входные данные и выходные данные и нужно определить процесс обладающий определенными свойствами необходимыми для системы.			

§2.2 Существующие решения.

LinkedWeb[9]-это проект соединения информации в сети интернет, которая не была ранее связана по посредством нового подхода к понятию данные. Это проект о том, как уменьшить барьеры в сети интернет между уже имеющимися связанными данными. Более точным будет определение, что LinkedWeb – это система понятий и инструментов, которая для описания признанно лучшей практики для изучения, обмена и связи разрозненных данных, информации и знаний в Semantic Web[10].



Однако стоит заметить, что LinkedData – это прежде всего некоторая часть Semantic Web. Точнее можно сказать, что данные в Semantic Web должны быть Linked, но при этом они совершенно не говорят о консистентности и алгоритме построения Semantic Structure. А значит саму задачу семантизации данных проект не решает, он лишь определяет протокол и алгоритм связи. А так же пытается построить Semantic Web из уже имеющихся семантических элементов, таких как некоторые части Wikipedia. Однако, в данной работе одним из ключевых элементов является использование социальной сети для построения семантической, конечно работа в данном направлении у LinkedWeb ведется, но стоит заметить, что LinkedData Project в данный момент не работает с отношением бизнес-социальная сеть:

Business with Linked Data

- What is the business value of Linked Data? [\[11\]](#)

Оставляя этот вопрос без ответа, хотя исходя из современных тенденций развития, данное направление будет одним из важнейших. Именно ему и уделяется ключевое внимание в данной работе.

§2.3 Японский подход к решению проблемы.

По мнению японского специалиста в области менеджмента Х. Йосихара [\[12\]](#), японский стиль управления характеризуется следующими особенностями:

1. Гарантия занятости и создание обстановки доверительности. Подобные гарантии, обеспечиваемые системой пожизненного найма, способствуют стабилизации трудовых ресурсов и уменьшают текучесть кадров. Стабильность служит стимулом для рабочих и служащих, она укрепляет чувство корпоративной общности, гармонизирует отношения рядовых сотрудников с руководством. Освободившись от давящей угрозы увольнения и имея реальную возможность для продвижения по вертикали, рабочие получают мотивацию для укрепления чувства общности с компанией. Стабильность также способствует улучшению взаимоотношений между работниками управленческого уровня и рядовыми рабочими, что, по мнению японцев, является залогом улучшения деятельности компании.
2. Гласность и ценности корпорации. В условиях, когда управленцы всех уровней и рядовые сотрудники начинают пользоваться общей базой информации о политике и деятельности фирмы, развивается атмосфера участия и общей ответственности, что улучшает взаимодействие работников и повышает производительность. Японская система управления старается способствовать унификации восприятия корпоративных ценностей, таких, как приоритет качественного обслуживания, услуг для потребителя, сотрудничество рабочих с администрацией, сотрудничество и взаимодействие отделов.
3. Управление, основанное на информации. Особое значение придается сбору данных и их систематическому использованию для повышения экономической эффективности производства и качественных характеристик продукции.
4. Управление, ориентированное на качество. Президенты фирм и управляющие компаний на японских предприятиях указывают на необходимость контроля качества. Главной заботой при управлении производственным процессом является получение точных данных о качестве.
5. Постоянное присутствие руководства на производстве. Для содействия решению проблем по мере их возникновения японцы зачастую размещают управляющий персонал прямо в производственных помещениях. По мере разрешения каждой проблемы вносятся небольшие нововведения, что приводит к накоплению дополнительных новшеств. В Японии для содействия дополнительным нововведениям широко используется система новаторских предложений и кружки качества.

6. Поддержание чистоты и порядка. Одним из существенных факторов высокого качества японских товаров являются чистота и порядок на производстве. Руководители японских предприятий стараются установить такой порядок, который может служить гарантией качества продукции и способен повысить производительность благодаря чистоте и порядку. Современные методы управления сложились в Японии в условиях послевоенной разрухи, которая поставила перед руководителями задачу восстановления социальной, политической и экономической жизни страны. Под влиянием американской оккупационной администрации будущие японские менеджеры познакомились с идеологией и методами управления бизнесом, распространенными в США. Это не означает, однако, что до 1945 года в Японии не существовало эффективной системы управления производством: ещё 5 мая 1932 года К. Мацусита, основатель компании «Мацусита Денки», которого в Японии называют «всесильным волшебником менеджмента» и «основателем вероучения об управлении», произнёс перед своими служащими речь, в которой отметил, что основным предназначением производителя является преодоление бедности.

Таким образом, мы можем видеть, что во многом проблема взаимодействия бизнеса и социальных сетей видоизменена для Японии в следствии пункта 1, пункта 5 и пункта 2. Во много наше приращение к решению задачи основано на схожих идея, но прописанных в контексте Европейского подхода и менталитета

§2.4 Почему применение семиотики действительно критично в данном контексте.

Прежде всего рассмотрим классический пример, который уже плотно вошел в лексикон современного пользователя интернета.

Рассмотрим английское существительное “friend” и глагол “to friend”. С точки зрения человека далекого от лингвистики и семиотики значение глагола to friend очевидным образом вытекает из значения существительного “friend” - друг, однако на самом деле, это оказывается неверным утверждением, хотя и кажущимся очевидным на первый взгляд(в реальности глагол “дружить” в английском языке передается конструкциями “to be a friend” и “to make a friend”).

Если мы обратимся к истории слова friend, то обнаружим, что глагол to friend является архаизмом и не применяется в английском языке с 15го века. Однако, с развитием социальных сетей этот глагол снова входит в оборот, но уже в совершенно ином качестве! В данном контексте, он обозначает добавить человека в список контактов, а вовсе не стать его другом, как может показаться на первый взгляд. Самая динамичная и быстро развивающаяся

социальная сеть в мире – Tweeter уже осознала данный факт и приняла соответствующие меры. Вместо популярного Friend List они используют Followers, что меняет дело, формируя определенное отношение к данной группе. То есть в данном случае, можно говорить, что глагол “to friend” семантически ближе к существительному follower, чем к noun friend.

В данном случае становится очевидным факт неверной интерпретации очень внешне схожих знаков, при этом имеющих совершенно различные денотаты, не говоря уже о коннотатах. И их смешение является грубым нарушением политики безопасности, которых стоит избегать при конструировании системы. Потому как оно формирует мнение о том, что список контактов социальной сети является “друзьями пользователя”, а значит влияет на восприятие знаковой системы в целом и на уровень доверия и открытости информации о пользователе в целом.

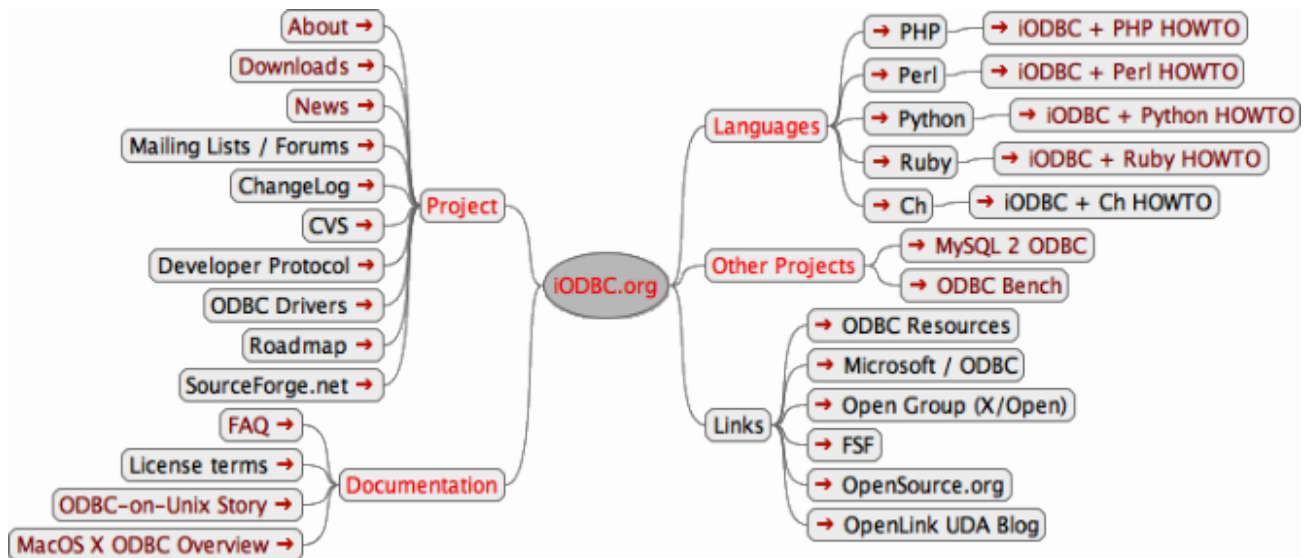


А значит, что правильная коррекция политики безопасности интранета и социальных сетей во всемирной паутине не может быть произведена без глубокого анализа знаковой системы. В данном случае легко видеть, что внешнее сходство знаков может привести к ложным выводам, даже в таком казалось бы простом случае. Тогда в свою очередь в более сложном знаковом произведении ошибки подобного рода могут привести к краху или

неработоспособности системы. Именно поэтому мы не можем обойти стороной анализ системы процессинга социальных сетей с точки зрения семиотики и знаковых систем.

Глава 3

§3.1 JDBC-ODBC аналогия.

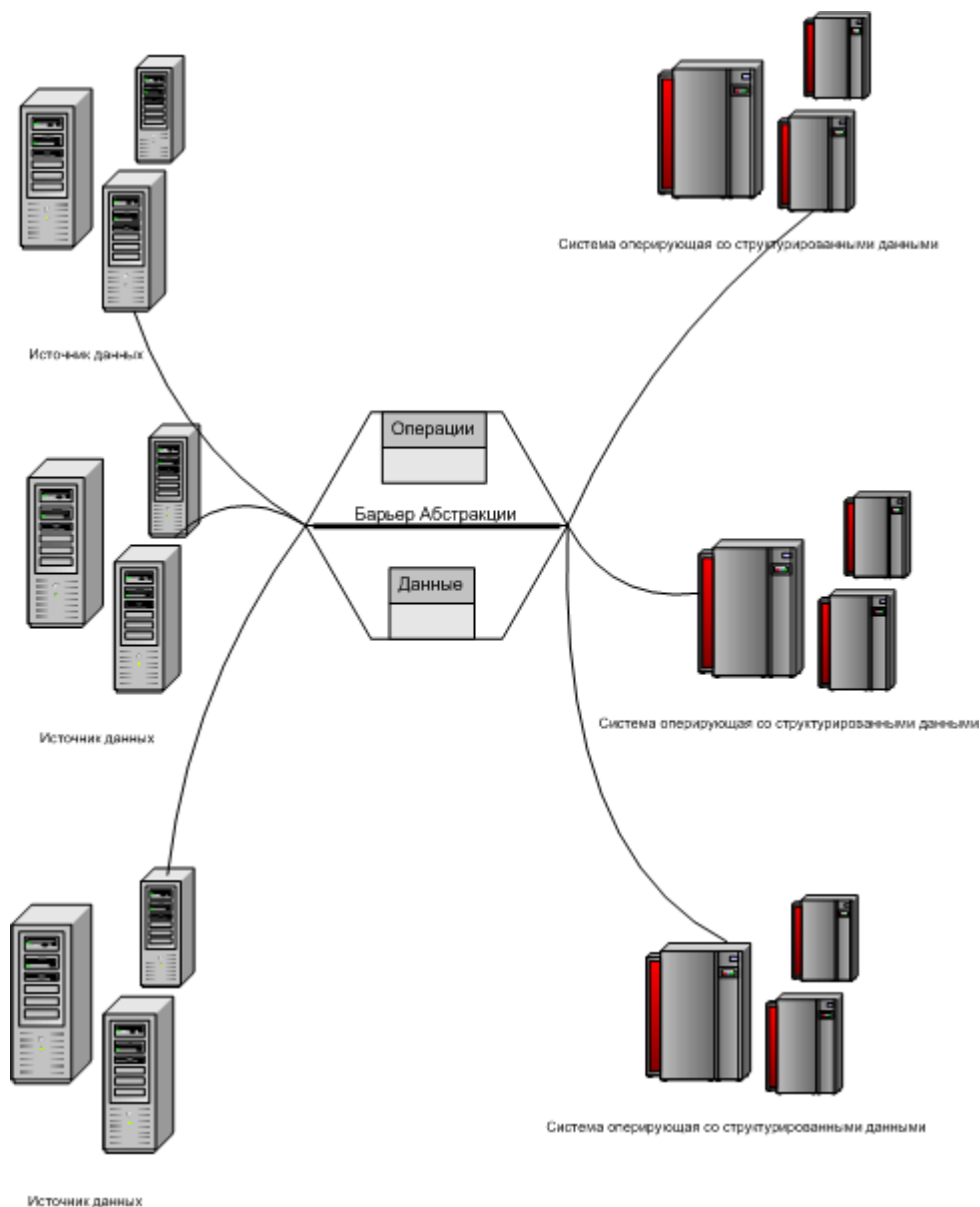


Технология ODBC[13] - это программный интерфейс (API) доступа к базам данных, позволяющий единообразно работать с разными источниками данных, абстрагируясь от особенностей взаимодействия в каждом конкретном случае.

Другими словами это есть некоторый мост между данными и агентами, использующими эти данные для своей работы. В качестве агентов могут выступать совершенно разнообразные системы, в том числе, и например, человек сидящий за терминалом.

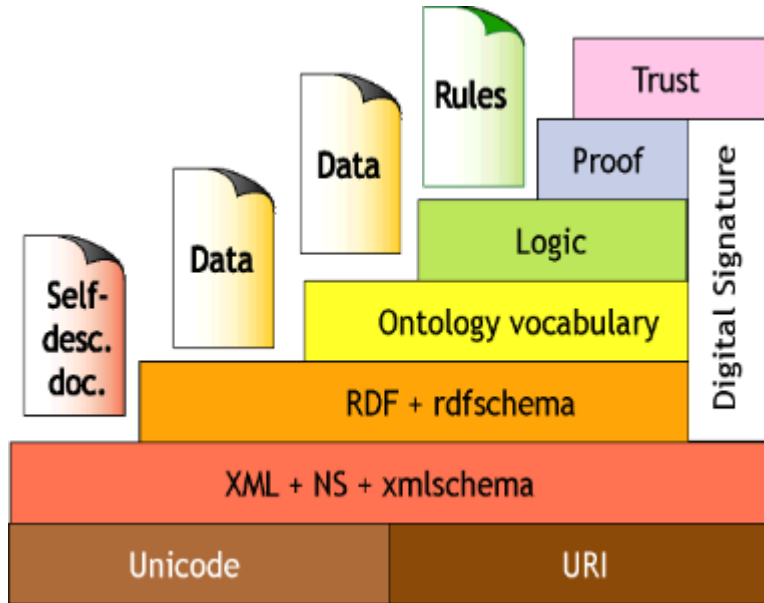
Теперь немного расширим понятие базы данных[14]: БД является представленная в объективной форме совокупность самостоятельных материалов (статей, расчетов, нормативных актов, судебных решений и иных подобных материалов), систематизированных таким образом, чтобы эти материалы могли быть найдены и обработаны с помощью электронной вычислительной машины (Гражданский кодекс РФ, ст. 1260). Здесь мы будем понимать ее именно в этом смысле, а не как конкретную имплементацию от различных вендоров, таких как Microsoft SQL Server, IBM DB 2, PostgreSQL, etc

Теперь становится понятным, что ODBC – это классический барьер абстракции[15] разделяющий модули системы. Для того, чтобы завершить аналогию совершим переход-подстановку:



Таким образом мы перешли от ODBC к классическому барьеру абстракции вида “мост”[\[16\]](#), который мы собираемся построить в нашей задаче. Стоит заметить важную особенность, что сами по себе данные из различных источников кажутся не связанными, на деле, если рассматривать их как систему, то они могут оказаться связанными. В связи с этим, необходимо обозначить некоторое концептуальное разделение между ними, которое бы с другой стороны позволяло им взаимодействовать и обладало некоторой гибкостью и стабильностью[\[17\]](#).

§3.2 Semantic Web.



The Semantic Web – развитие идей глобальной паутины, в котором ей приписывается некоторый смысл, иначе говоря семантика, которая позволяет [\[18\]](#) машинам обрабатывать текст. Иначе говоря в этом случае обработчик работает не с сырыми данными(row data), а с некоторой четко выделенной структурой, которую мы, в след за специалистами, будем именовать Semantic structure[\[19\]](#).

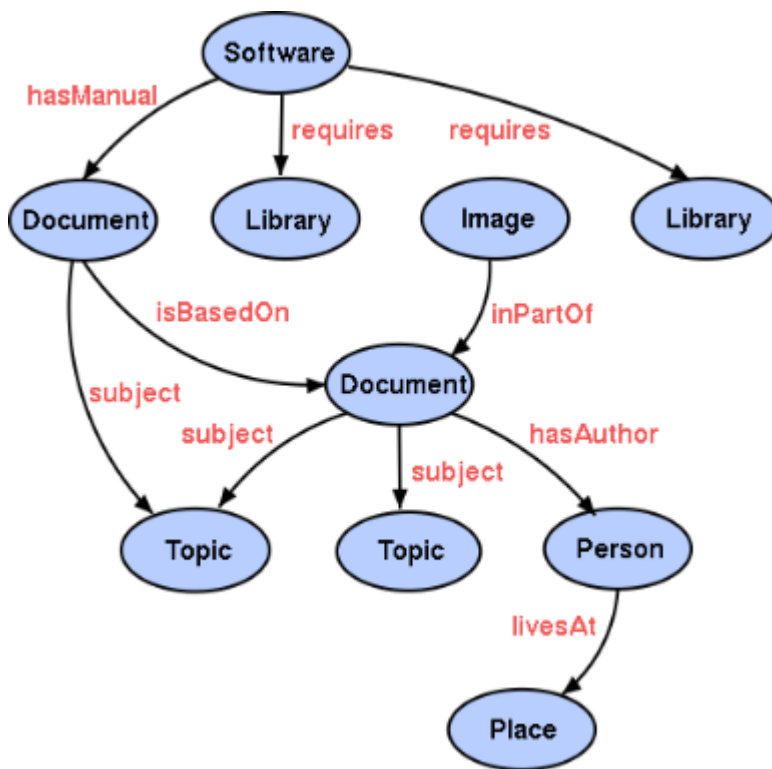
В основе Semantic Web лежит несколько фундаментальных понятий Ontology Representation, Knowledge Systems and Self-Description

Knowledge Systems[\[20\]](#) в данном контексте означает, что система способна извлекать факты\сведения и информацию на основе заложенных в нее данных.(классический пример с деревом вывода)

Ontology Representation[\[21\]](#) здесь означает, что у нас есть единый метод представления и обращения к данным

Self-Description[\[22\]](#) здесь означает, что система в состоянии анализировать собственное поведение. Иначе говоря, система может не только выдать значения определенного предиката, но и продемонстрировать вывод определенного выражения(why, what).

В частности подобный подход также используется в проектировании разработке ПО.



Как можно заметить, Semantic Web и все его элементы восходят к такому понятию как метаданные[23].

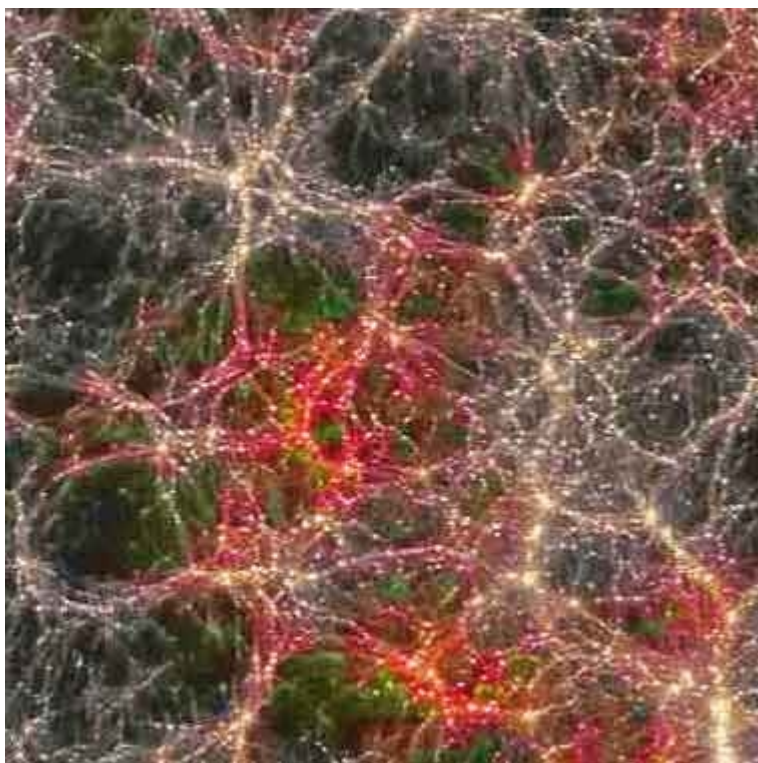
Сама идея идея Semantic Web состоит в построении подобной структуры, но тут мы приходим к новой проблеме:

The social web or the semantic web ?[24]

Так же согласно многим другим авторам[25], задача о построении семантической структуры в общем виде выглядит аналогично дилемме о “Курице и Яйце”(с технической точки зрения эквиволентной задаче о раскрытии компилятора[26])

Которая в нашем случае разрешается с помощью социального элемента, который даст достаточный объем обучающей выборки, для построения метрик и структур над заданным множеством. Таким образом выбор стоит не между social web or semantic web, в том смысле будет ли глобальная паутина быть значимой и семантически ориентированной на обработку человеком или ориентированной на обработку машиной. А скорее центральным будет их взаимное сосуществование и взаимодействие, как мы это уже видели ранее. Во

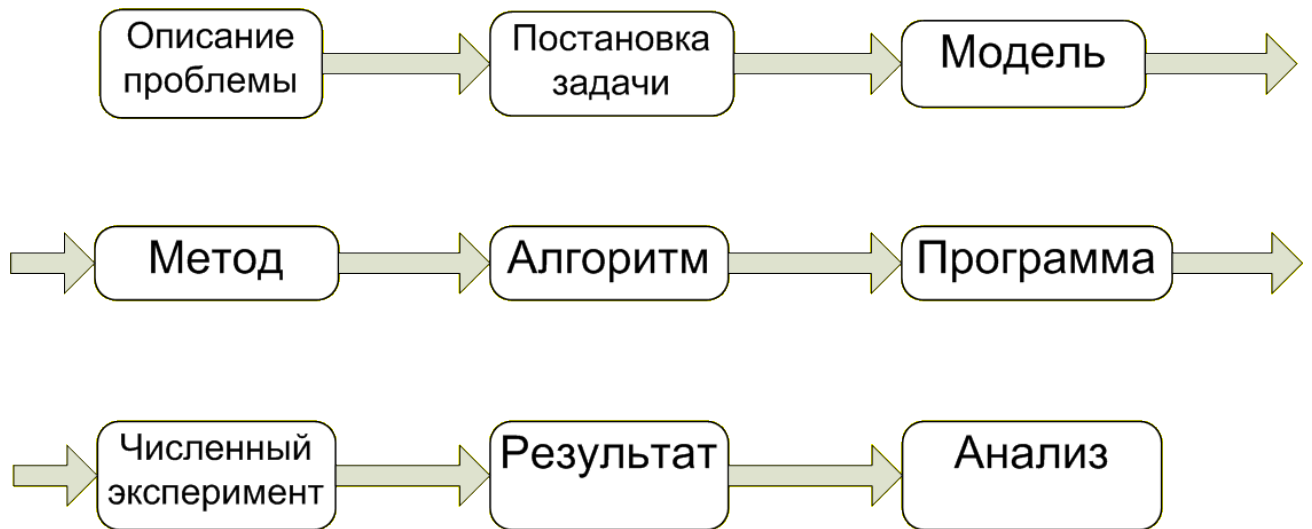
многим мы не можем приписать значения определенным параметрам или системам, не опираясь на эвристические показатели или результаты экспертов, как это сделано в экспертных системах, или во многом культивируется известными социальными тематическими медиа, такими как например: last.fm[27]. Когда введение метрики является ключевым в существовании сервиса, а его введение невозможно без эвристических или экспертных статистик и алгоритмов, в данном случае, нельзя обойтись без социальной структуры, которая как раз и позволяет сделать “надстройку”-метаданные и определяющие семантическую структуру системы.[28]



Данная иллюстрация показывает степень значимости данных для человека и связи между ними. Чем более актуальными являются данными, тем более горячими они изображены. С другой стороны, температура (“степень нагретости”) может так же передаваться через социальные связи, которые так же воспринимаются как часть структуры. В такой интерпретации, мы имеем наглядное изображение неравнозначности данных в глобальной паутине, которая должна быть учтена.

Глава 4. Модель.

§4.1 Схема



Прежде всего в основе нашей модели лежит вышеуказанная схема.

Приведем краткое описание нашей интерпретации данной схемы:

Описание проблемы — это неформальное представление задачи, предпосылка, исторические сведения, вся та информация, которую мы подчерпнули перед решением проблемы.

Постановка задачи — это формализация той информации, которой мы обладаем по данной задаче. Это так же постановка критериев оценивания. Входных и выходных данных.

Модель является представлением объекта, системы или понятия в форме, отличной от реальной, но приближенной к алгоритмическому описанию, включающей и набор данных, характеризующих свойства системы и динамику их изменения со временем¹

Под методом здесь мы будем понимать общую идею или паттерн², которую можно использовать для целого класса задач(пример, идеология в алгоритмах «разделяй-и-властвуй»)

Алгоритм здесь это конкретное использование элементов модели, инструментария нашего метода в заданной последовательности

Программа — конкретная имплементация нашего алгоритма

Численный эксперимент — прогонка нашей программы на конкретной задаче и\или наборе тестов.

Результат — полученные выходные данные

Анализ — изучение возможностей дальней исследований.

1 [Ножнов В.А.](#) Модель учебного курса. //Сборник трудов научной конференции ИТО-2009.

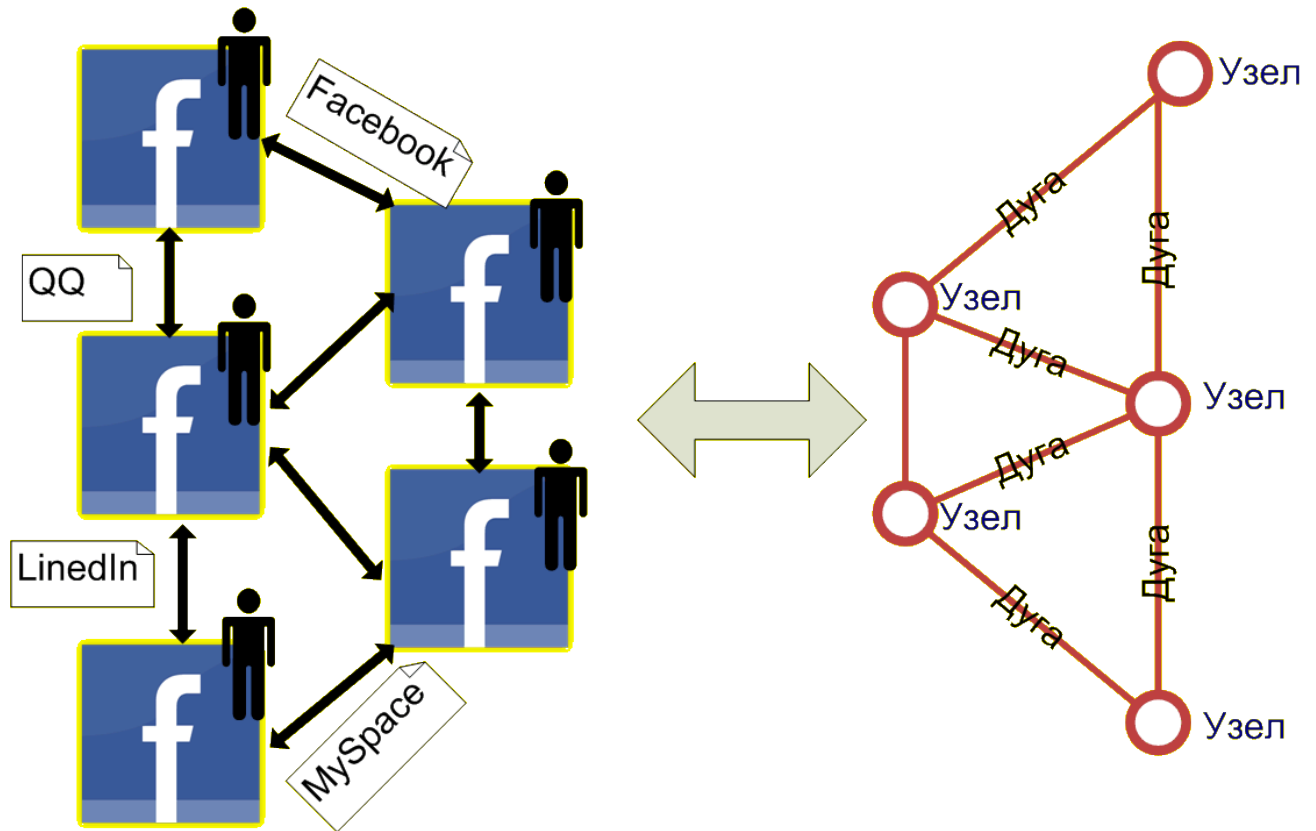
2 Gamma, E, Helm, R, Johnson, R, Vlissides, J: *Design Patterns*, page 18. Addison-Wesley, 1995

§4.2 Элементы модели.

Прежде всего нужно отметить, что модель должна выделять одни свойства и игнорировать другие. Значит, что нам нужно оговорить, что же мы выделяем и в каком виде.

Начнем с первого элемента:

Сетей.



Мы будем полагать, что социальная сеть и интранет представимы в виде ориентированного графа.

Для социальных сетей этот факт очевиден, а для интранет-сетей он требует обоснования и является ключевым для всей работы. Его обоснование будет приведено в главе (программа и алгоритм)

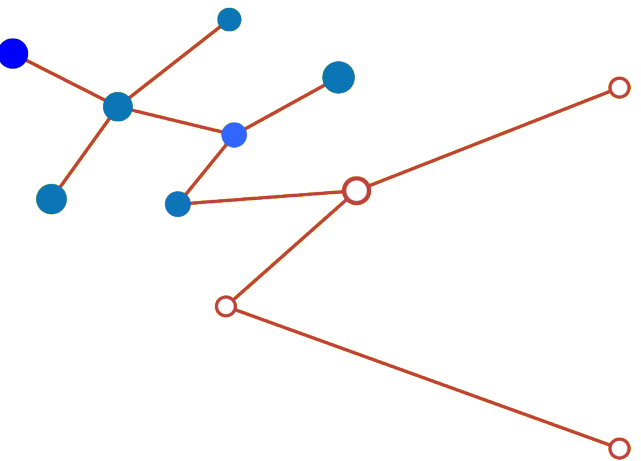
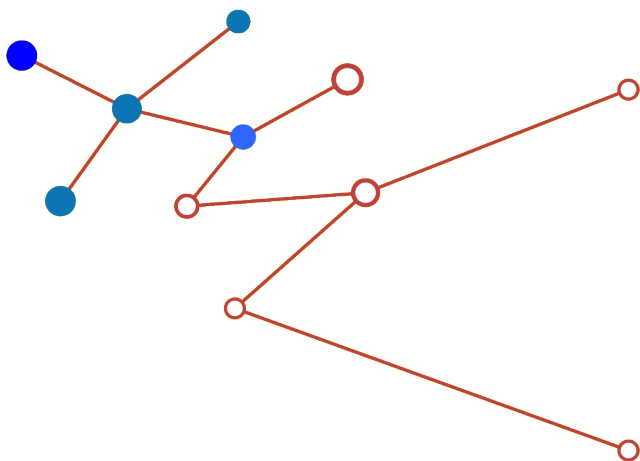
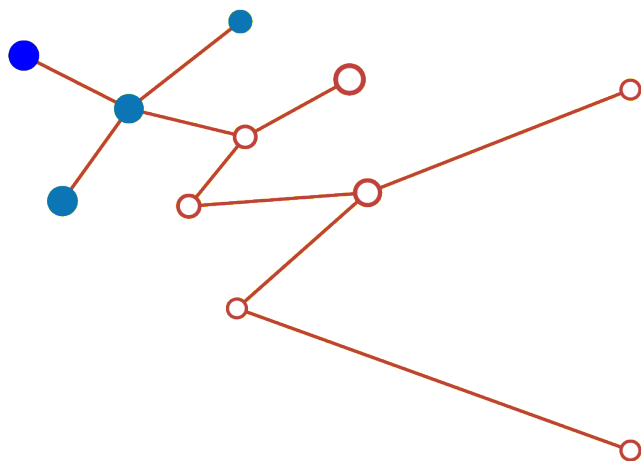
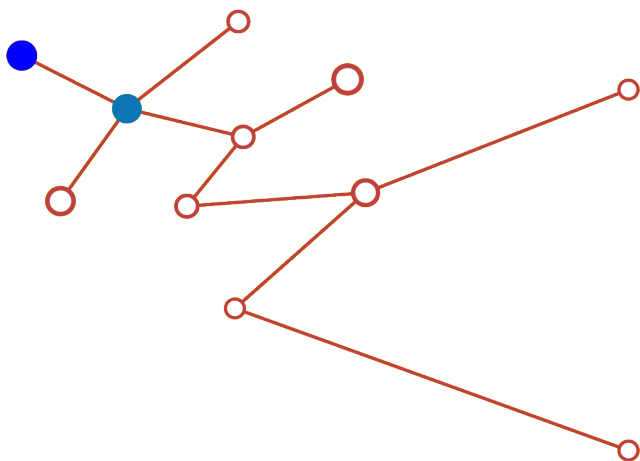
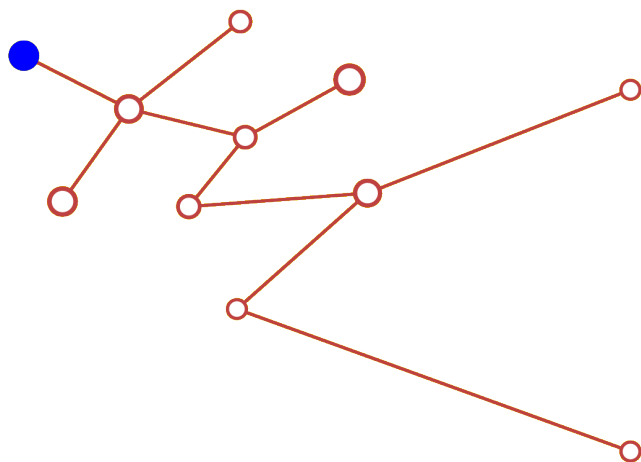
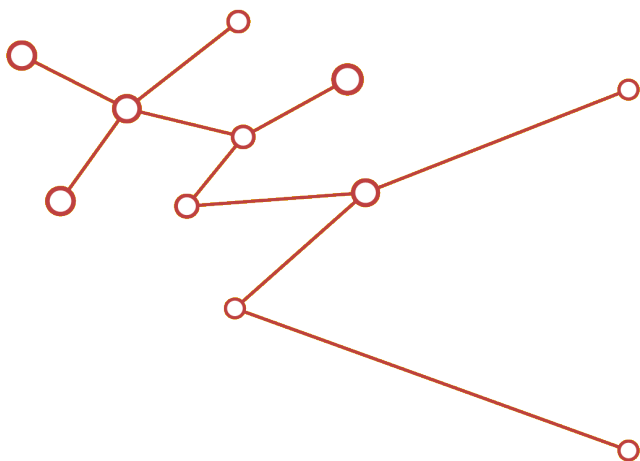
Фактически из данного предположения автоматически следуют все возможные операции над элементами модели, так как они являются графами в классическом смысле. То есть, мы можем применять все классические графовые операции и алгоритмы.

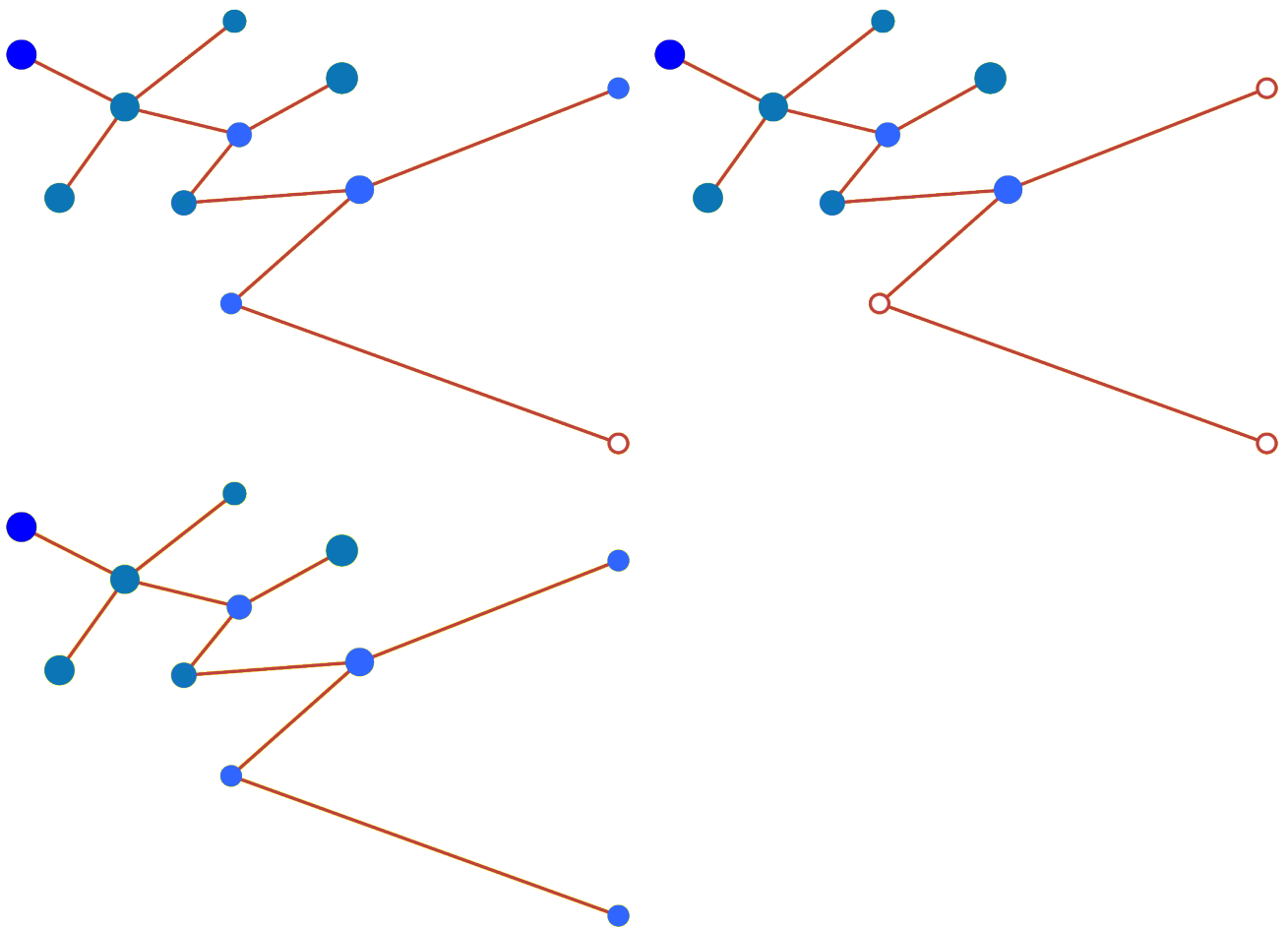
Операции над сетями:

Обходы графа — поиск в глубину и поиск в ширину

Поиск кратчайшего пути - алгоритм Дейкстры, алгоритм Флойда

Потоки в графах — Форд-Фалкерсон, Беллман-Форд, алгоритм проталкивания предпотока и другие, для наглядности приведем один из них:





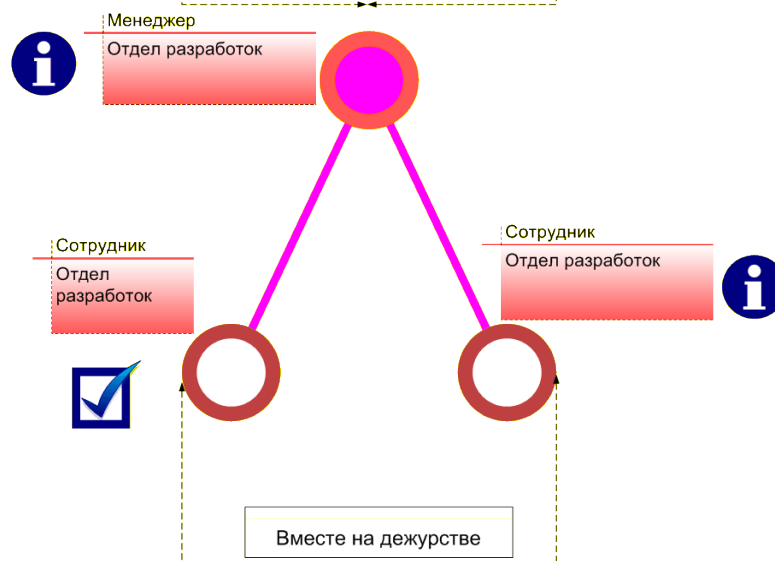
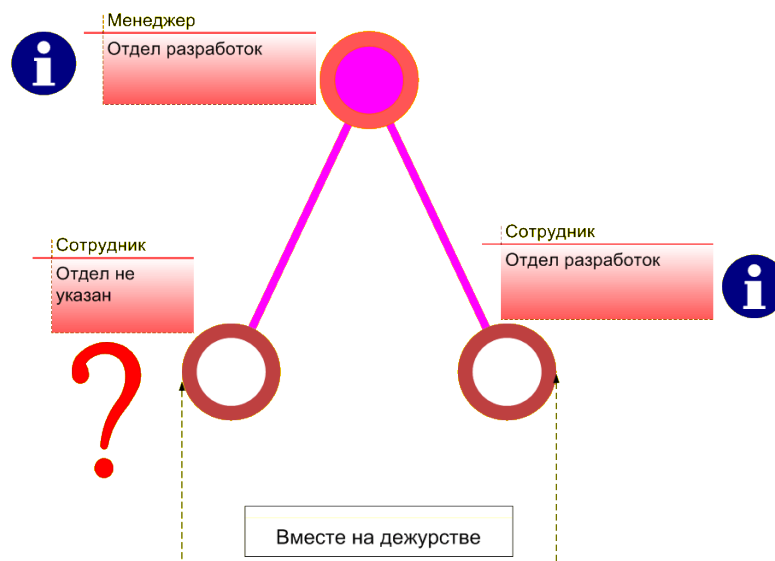
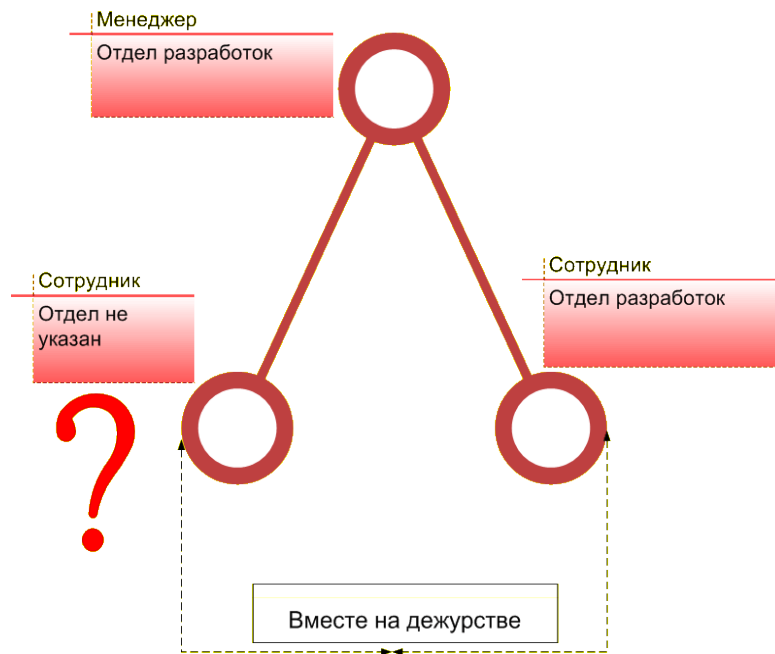
Данный алгоритм поиска в ширину может применяться, например, в следующей задаче:

- 1) Пусть задан граф, ребра которого имеют единичный вес(например, когда граф задает карту лабиринта)
- 2) Необходимо найти кратчайший путь из вершины А в Б(например, как показано на рисунке из вершины с левого верхнего угла в правый нижний)

Метаструктура сети

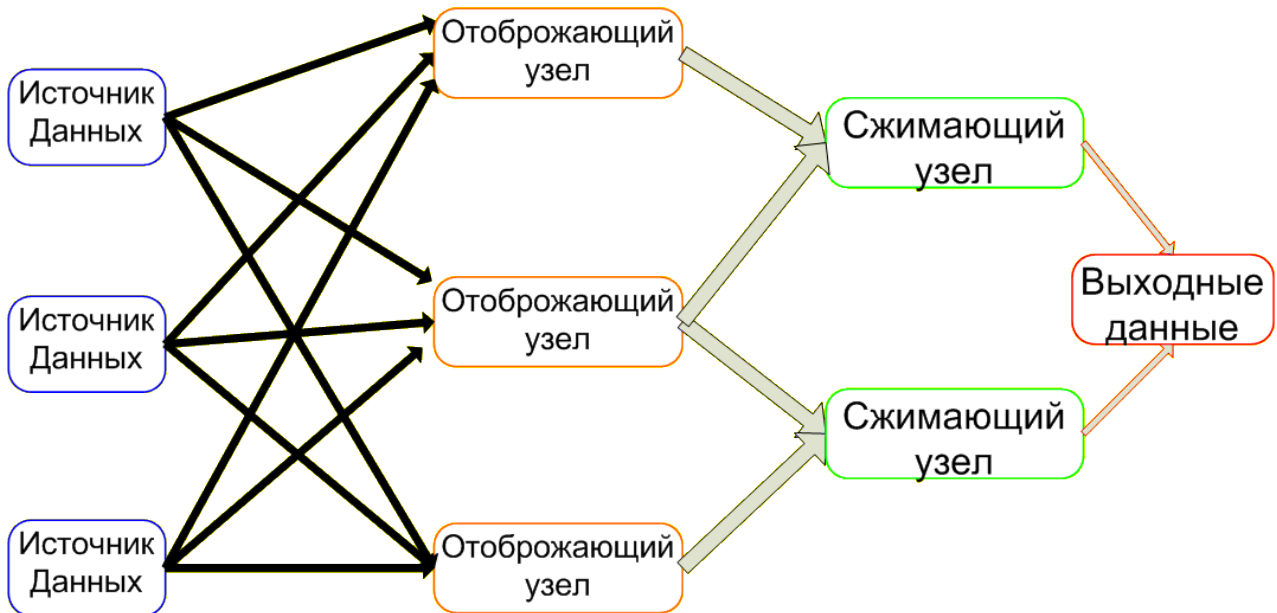
Здесь под метаструктурой мы будем понимать наличие структуры, которая бы описывала саму структуру нашей модели(программы). Это немного нам для следующей цели: система должна быть способна не только производить вычисления, но и быть способной отвечать на вопросы о своем поведении, то есть отвечать на вопросы как и почему? Быть способной делать простейшие выводы о свойствах объектов, даже если информация о них не был явно заложена в систему, тут было бы уместно привести еще один пример.

Пусть даны три работника, один из которых менеджер и двое сотрудники. Причем известно, что менеджер и один из сотрудников работают в отделе разработка и нужно ответить на вопрос: где работает третий сотрудник, если оба сотрудника находятся вместе на дежурстве.



Глава 5. Алгоритм и метод.

Прежде всего стоит отметить, что мы будем пользоваться общепризнанной терминологией и паттерном, чтобы избежать излишней детализации при работе над нашей основной задачей. В основу нашего метода был положен знаменитый паттерн MapReduce[29]:



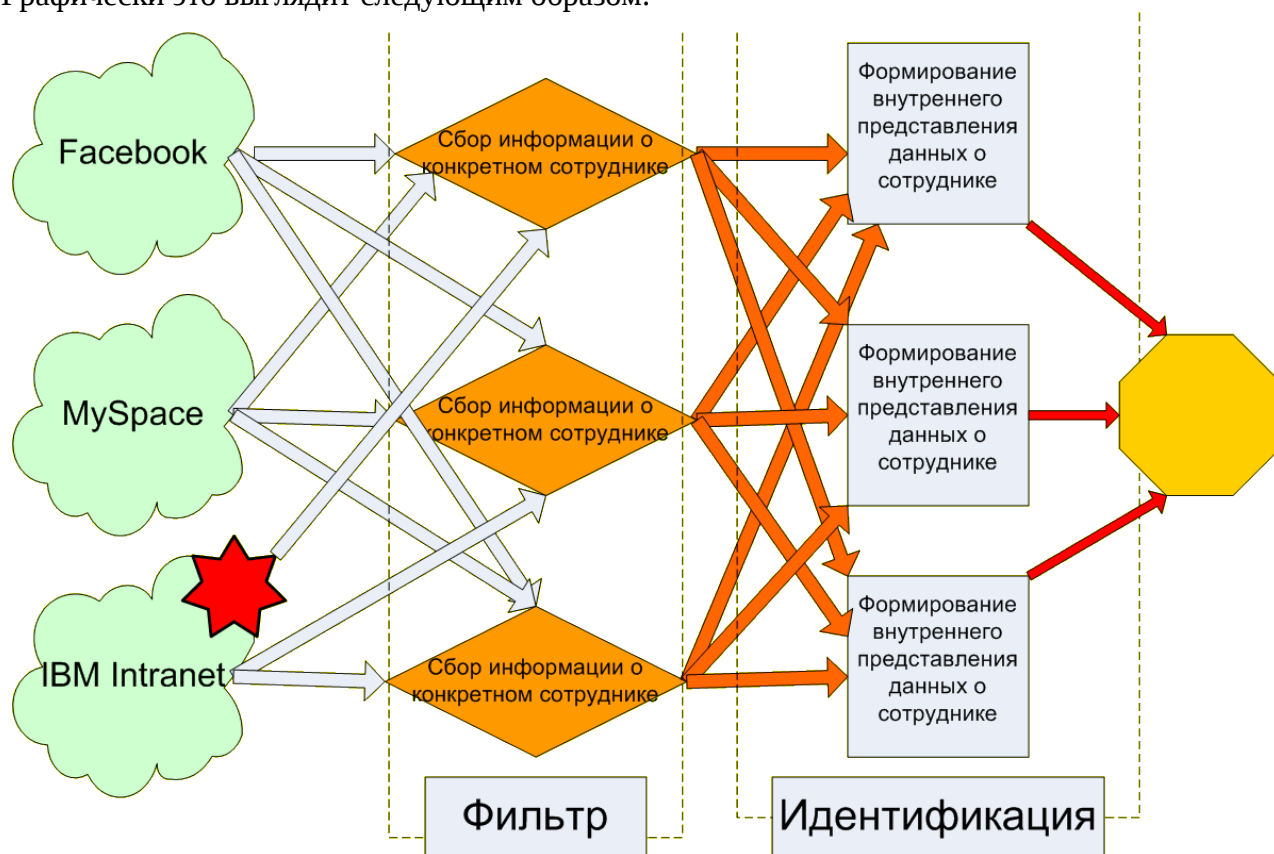
В нашей интерпретации метод MapReduce стоит понимать в следующем смысле:

В качестве источника данных выступают элементы модели — сети, ведь именно они обладают все содержащейся информацией. При выполнении отображения — Мар мы используем доступные над сетями операциями — обходами графов, поиском путей, etc.

Когда мы производим сжатие свою роль играет метаструктура сети, а так же все доступные призраки[30], что позволяет сделать максимальное полное собрание доступной информации о человеке.

Как результата мы получаем некоторую персонализированную единую базу, специальным образом заточенную под нашу задачу.

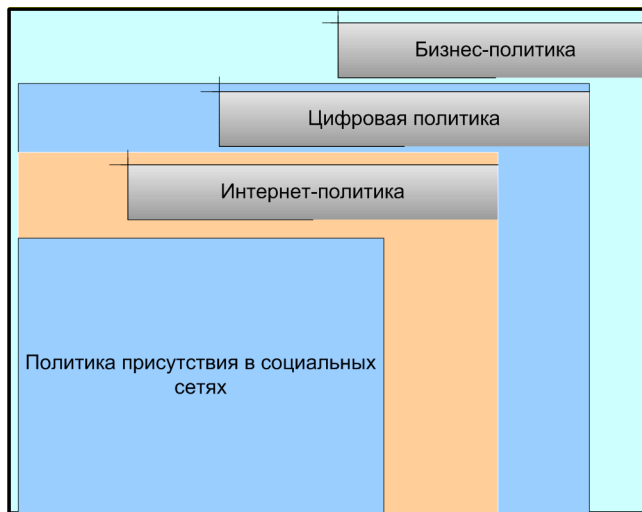
Графически это выглядит следующим образом:



Как можно видеть, IBM Intranet помечена здесь специальным символом — астерикс.

Это сделано умышленно, чтобы обратить внимание на принципиальную трудность построения данного алгоритма. В нашей модели интранет представим в таком же виде, что и сама соц. сеть(в виде графа, с графовыми операция и метаструктурой), однако данный факт оставался недоказанным. А значит, что первоочередной задачей для нас является показать принципиальную возможность выражения интранета в таком виде.

§5.1 Составляющие алгоритма.



Политика присутствия

Прежде всего это обозначение политики присутствия в социальной сети в рамках бизнес-политики.

Хотя данный аспект невероятно важен, он все же лежит немного вне пределов нашего доклада (потому как именно технических трудностей он не вызывает), поэтому ограничимся только обзором данного пункта.

Имеет смысл в качестве верхнего предела

видимости и доступности данных группу-представитель компании в данной социальной сети. Иначе говоря, вся информация, которую вы размещаете для публичного доступа в социальной сети, должна быть доступна группе. Иначе говоря, группа-представитель является самым внешним кругом видимости, что можно обосновать тем фактом, что человек даже в нерабочее время является лицом компании согласно современным европейским контрактам, значит существенно влияет на брендовые риски. А уменьшение брендовых рисков является задачей номер №1.

Создание метаструктуры

Данный вопрос довольно подробно обсуждался в главе 3.2, поэтому здесь мы ограничимся упоминанием того факта, что парадокс раскрутки компилятора решается с помощью построения семантической метрики из социальной (как, например, это сделано на Last.fm), в остальном аналогия совпадает с точностью до предоставляемого интерфейса.

Интеграция на основе открытой платформы

Данный пункт алгоритма является ключевым и для его разрешения, нам необходимо показать, что мы действительно можем обращаться с интранетом, так же как мы обращаемся с обычной сетью.

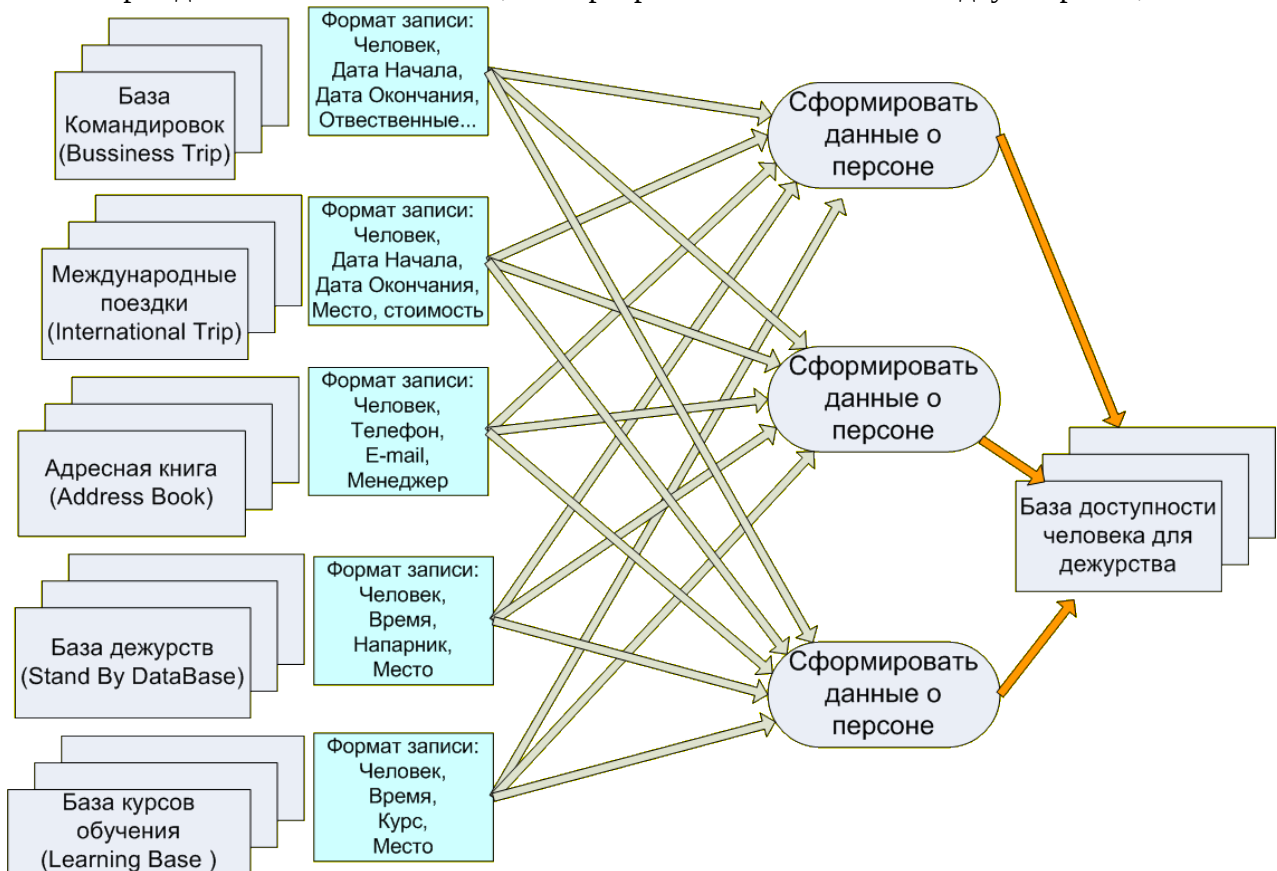
Конечно совершить обзор всех существующим систем документо-оборота фактически невозможно, но тут невероятно важно подчеркнуть следующий факт: практически все основные системы используемые в промышленности и бизнесе обладают предельно схожим функционалом, с точностью до интерфейса. Это значит, что функционал одной системы выразим через функционал другой системы.

Поэтому мы покажем, что для одной из наиболее популярных и крупных система-
документооборота мы можем явно построить такое преобразование и применить наш
алгоритм. А значит, для большинства современных сетей мы можем его применить, в силу
выразимости и сводимости интерфейсов.

Глава 6.

§6.1 Программа.

Прежде всего стоит отметить, что программа была написана в двух версиях, хотя обе

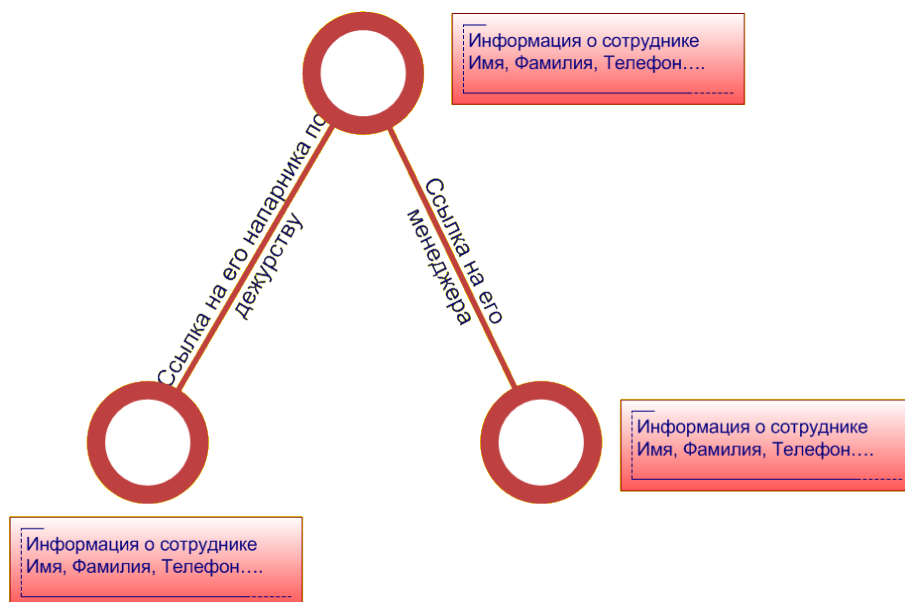


они укладываются в указанную выше схему. С небольшой конкретизацией функций Map и Reduce: Map здесь - это получение записи из базы данных о конкретном человеке, иначе говоря на одном из серверов IBM Russia System and Technology запускалось N агентов, где N число доступных исполнителей системы. Каждый агент собирал информацию о конкретном человеке и после сбора он отдавал эту информацию другим агентам которые заинтересованы в людях упомянутых в информации данного агента.

Пример, агент EclipseAgentBase (название агента-класса использованного в программе с использованием Lotus Java API) собирает все о сотруднике Иванове, в том числе и запись о том, что менеджер Иванова — Петров, и что Иванов дежурит с Сидоровым.

Значит наш EclipseAgentBase передаст эту информацию трем другим агентам ReduceAgent, которые отвечают за Иванова, Петрова и Сидорова, а те в свою очередь сформируют представление о конкретном человеке. Соответственно Reduce — это работа по формированию нашего внутреннего представления о человеке, которое фактически дает нам

его в виде инфо о сотруднике и связи внутри компании. На вышеуказанном примере, это выглядит следующим образом.



Таким образом мы приходим именно к нужному нам представлению данных, которое позволяло нам использовать наши метаструктуры и графовые алгоритмы, а еще что не менее важно — свойство замыкания(closure), что позволяет нам безболезненно состековать данные из различных источников.

Глава 7. Эксперимент

§7.1 Численный эксперимент

В основу эксперимента была положена уже указанная выше схема с пятью различными базами данных. Было использовано две различные схемы:

Первая, использование Lotus Domino Java API из Notes.jar файла, который поставляется вместе с Lotus Notes и Designer Professional .

Второй, использование ODBC-JDBC моста(от этой технологии подробнее написано в главе 3, параграфе 1).

Суть эксперимента состояла в том, чтобы распознать записи из 5ти баз: Командировок, Международных поездок, Адресной Книжки, Базы Дежурств и Базы курсов обучения. На выходе мы должны получить другую базу, с индивидуальной информацией по каждому сотруднику, чтобы быть способным отвечать на вопрос: может ли данный сотрудник выйти на дежурстве в указанный момент времени?

Приведем результаты вычислений:

Процент распознавания нашим отображением(Мар) доступных баз					
	Командировки	Междн. поездки	Адресная книга	База дежурств	Курсы обучения
ODBC-JDBC Bridge	87	91	94	89	86
Lotus Java API	91	93	95	91	89

Тут стоит сказать пару слов в защиту метода с мостом, во-первых это кроссплатформенность, обеспеченная наличием ODBC драйверов для большинства современных платформ. Во-вторых, переносимость на уровне кода, с точностью до коррекции структуры баз данных(а точнее соответствующих SQL запросов). В-третьих, недостаточная проработанность преобразования из документо-ориентированной базы данных в реляционную, в связи со сложностью ее изучения, но это скорее минус в сторону разработчика, а не самого метода.

В целом, среднее число нераспознанных записей у метода с мостом ~ 10%, а у метода с Lotus Java API ~8%. Причем, данные показатели явно улучшаемы, если уделить больше внимания проработанности структуре баз данных, что не было сделано в связи с большой занятостью их разработчиков и общей сложностью самой системы.

§7.2 Результат

Разработан протокол на примере системы IBM Lotus Domino в рамках автоматизации бизнес процессов (в IBM Russia System and Technology Laboratory)

Организовано два варианта кроссплатформенной репрезентации и передачи данных: XML(для моста), nsf (для Lotus Notes API).

Гипотеза о представлении не противоречит статистическим данным в пределах ошибки (10% для моста и 8% для Lotus Notes API)

Что фактически означает провомерность применения нашего алгоритма.

§7.3 Дальнейшее направление исследований

Анализ бизнес-процессов при проектировании системы разграничения прав и доступа, как модернизация пункта 1 нашего алгоритма — права и уровень доступа.

Обеспечение консистентности данных, как модернизация пункта 3 нашего алгоритма — интеграция через открытую платформу.

Необходимо разработать общий распределенный паттерн для интеграции обладающий свойством замыкания , как модернизация пунктов 2 и 3 — итеграции и создания подходящей распределенной метаструктуры.

Список литературы.

-
- [1] Succeeding through service innovation. White Paper. – Cambridge, IBM, 2008.
- [2] Беляев И. П., Капустян В. М., Рыков В. В. Знаковые структуры и бизнес процессы // Труды МФТИ. – 2009 - №.3.
- [3] Succeeding through service innovation. White Paper. – Cambridge, IBM, 2008.
- [4] Парамонов С.В. К вопросу о методике применения контент и факторного анализа в исторических исследованиях // Наука: первые шаги. – 2006 – №.1. – С. 34-43.
- [5] Рыков В. В. Обработка нечисловой информации. Управление знаниями. – М.: МФТИ, 2007.
- [6] [Social Networking Competition](http://about.ovum.com/sncomp/) <http://about.ovum.com/sncomp/>
- [7] Арнольд В. И. Жёсткие и мягкие математические модели. — М.: МЦНМО, 2004, С. 27-29
- [8] Новиков А.М., Новиков Д.А. "Методология". М.: Синтер, 2007.- 614-617с.
- [9] Christian Bizer, Tom Heath and Tim Berners-Lee (in press). **Linked Data - The Story So Far**. International Journal on Semantic Web and Information Systems, Special Issue on Linked Data.
- [10] [Linked Data - Connect Distributed Data across the Web](http://linkeddata.org/home) <http://linkeddata.org/home>
- [11] [Linked Data - Connect Distributed Data across the Web](http://linkeddata.org/faq) <http://linkeddata.org/faq>
- [12] Х.Р. Ёсихара Управление производством за рубежом // Интерсоциоинформ: Маркетинг Успеха – 2000 - №6.
- [13] [Википедия](http://en.wikipedia.org/wiki/Open_Database_Connectivity) http://en.wikipedia.org/wiki/Open_Database_Connectivity
- [14] [Википедия](http://ru.wikipedia.org/wiki/База_данных) http://ru.wikipedia.org/wiki/База_данных
- [15] [Hal Abelson's, Jerry Sussman's and Julie Sussman's Structure and Interpretation of Computer Programs](http://mitpress.mit.edu/sicp/full-text/book/book-Z-H-14.html#%_sec_2.1.2) http://mitpress.mit.edu/sicp/full-text/book/book-Z-H-14.html#%_sec_2.1.2
- [16] [Википедия](http://en.wikipedia.org/wiki/Bridge_pattern) http://en.wikipedia.org/wiki/Bridge_pattern
- [17] Gamma, E, Helm, R, Johnson, R, Vlissides, J: *Design Patterns*, page 151. Addison-Wesley, 1995
- [18] Berners-Lee, Tim; James Hendler and Ora Lassila (May 17, 2001). ["The Semantic Web"](http://www.sciam.com/article.cfm?id=the-semantic-web&print=true). *Scientific American Magazine*. <http://www.sciam.com/article.cfm?id=the-semantic-web&print=true>. Retrieved March 26, 2008.
- [19] [Википедия](http://en.wikipedia.org/wiki/Semantic_Web) http://en.wikipedia.org/wiki/Semantic_Web
- [20] [Википедия](http://en.wikipedia.org/wiki/Knowledge_representation) http://en.wikipedia.org/wiki/Knowledge_representation

[21] [Википедия](http://en.wikipedia.org/wiki/Web_Ontology_Language) http://en.wikipedia.org/wiki/Web_Ontology_Language

[22] [Википедия](http://en.wikipedia.org/wiki/Autoepistemic_logic) http://en.wikipedia.org/wiki/Autoepistemic_logic

[23] [Library of Congress Washington DC on metadata: http://www.loc.gov/standards/metadata.html](http://www.loc.gov/standards/metadata.html)

[24] http://opengardensblog.futuretext.com/archives/2007/10/beyond_web_20_t.html

[25] [Ajit Jaokar Beyond Web 2.0: The social web or the semantic web ?](http://www.webdesignfromscratch.com/blog/future-social-web-experience/)

<http://www.webdesignfromscratch.com/blog/future-social-web-experience/>

[26] *Альфред Ахо, Рави Сети, Джеффри Ульман* Раскрутка // Компиляторы: принципы, технологии и инструменты = *Compilers: Principles, Techniques, and Tools*. — М.: Вильямс, 2003. — С. 681—684. — 768 с. — ISBN 5-8459-0189-8

[27] [Last.fm music service](http://www.last.fm/) <http://www.last.fm/>

[28] Классической иллюстрацией используемой в Semantic Web неоднозначности данных является heat-map, Ben Hunt future-social-web-experience: <http://www.webdesignfromscratch.com/blog/future-social-web-experience/>

[29] ["MapReduce: Simplified Data Processing on Large Clusters"](#), by Jeffrey Dean and Sanjay Ghemawat; from [Google Labs](#)

[30] Н.Н. Непейвода: Стили и методы программирования
<http://www.intuit.ru/departments/se/progstyles/3/footnote.3.1.htm>