

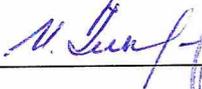
МИНОБРНАУКИ РОССИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
«КОЛЬСКИЙ НАУЧНЫЙ ЦЕНТР РОССИЙСКОЙ АКАДЕМИИ НАУК»
(ФИЦ КНЦ РАН)

УТВЕРЖДАЮ

Начальник управления аспирантуры и магистратуры
ФИЦ КНЦ РАН

к.г.-м.н., доцент И.В. Чикирёв




ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

По дисциплине Б1.О.02 Научно-исследовательский семинар (Методология научных исследований и разработок)
указывается цикл (раздел) ОП, к которому относится дисциплина, название дисциплины

для направления подготовки (специальности) 09.04.02 Информационные системы и технологии
код и наименование направления подготовки (специальности)

направленность программы (профиль) Информационные системы предприятий и учреждений
наименование профиля /специализаций/образовательной программы

Квалификация выпускника, уровень подготовки

Магистр

(указывается квалификация (степень) выпускника в соответствии с ФГОС ВО)

Апатиты

2020

Лист согласования

1. Разработчик:

доцент
должность

УАиМ


подпись

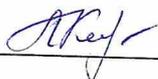
А.В.. Вицентий
И.О. Фамилия

2. Фонд оценочных средств рассмотрен и одобрен на заседании учебно-методической комиссии управления аспирантуры и магистратуры 29 июня 2020 г., протокол № 02.

Председатель УМК УАиМ

29.06.2020
дата

подпись



Л.Д. Кириллова
И.О.Фамилия

Лист переутверждения

Фонд оценочных средств переутвержден на 2021/2022 учебный год без изменений и дополнений.

Председатель УМК УАиМ Л.Д. Кириллова Л.Д. Кириллова

Основание: протокол № 2 от «29» июня 2021 г.

Фонд оценочных средств переутвержден на _____ / _____ учебный год без изменений и дополнений.

Председатель УМК УАиМ _____ Л.Д. Кириллова

Основание: протокол № _____ от « _____ » _____ 20 _____ г

Фонд оценочных средств переутвержден на _____ / _____ учебный год без изменений и дополнений.

Председатель УМК УАиМ _____ Л.Д. Кириллова

Основание: протокол № _____ от « _____ » _____ 20 _____ г

Фонд оценочных средств переутвержден на _____ / _____ учебный год без изменений и дополнений.

Председатель УМК УАиМ _____ Л.Д. Кириллова

Основание: протокол № _____ от « _____ » _____ 20 _____ г

Фонд оценочных средств переутвержден на _____ / _____ учебный год без изменений и дополнений.

Председатель УМК УАиМ _____ Л.Д. Кириллова

Основание: протокол № _____ от « _____ » _____ 20 _____ г

Лист изменений, вносимых в ФОС по дисциплине «Научно-исследовательский семинар (Методология научных исследований и разработок)»

В фонд оценочных средств вносятся следующие изменения и дополнения:

1. _____

2. _____

3. _____

Дополнения и изменения внесены и одобрены на заседании учебно-методической комиссии управления аспирантуры и магистратуры ФИЦ КНЦ РАН

от «____» _____ г., протокол № _____.

Председатель УМК УАиМ _____ Л.Д. Кириллова

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Общие сведения

1.	Управление	Аспирантуры и магистратуры
2.	Направление подготовки	09.04.02 Информационные системы и технологии Профиль – «Информационные системы предприятий и учреждений»
3.	Дисциплина (модуль)	Б1.О.02 Научно-исследовательский семинар (Методология научных исследований и разработок)
4.	Количество этапов формирования компетенций (ДЕ, разделов, тем и т.д.)	2

Перечень компетенций:

- | |
|---|
| <ul style="list-style-type: none">• Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий (УК-1)• Способен применять на практике новые научные принципы и методы исследований (ОПК-4) |
|---|

Критерии и показатели оценивания компетенций на различных этапах их формирования

Этап формирования компетенции (разделы, темы дисциплины)	Формируемая компетенция	Критерии и показатели оценивания компетенций			Формы контроля сформированности компетенций
		Знать:	Уметь:	Владеть:	
1. Основы научно-исследовательской деятельности. Основные понятия методологии научных исследований и разработок. История и тенденции развития методологии научно-исследовательской деятельности. Законодательные основы научных исследований и разработок.	УК-1	принципы сбора, отбора и обобщения информации	соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности	работы с информационными источниками, опыт научного поиска, создания научных текстов	Оценка результатов практического занятия, контрольная работа.
	ОПК-4	новые научные принципы и методы исследований	применять на практике новые научные принципы и методы исследований	применения новых научных принципов и методов исследования для решения профессиональных задач	
2. Методологические основы научных исследований и разработок. Понятие о методе и методологии исследования. Уровни методологии. Универсалии науки. Типология методов научных исследований. Логика научного исследования. Алгоритм научного исследования. Методологические основы теоретических научных исследований и разработок.	УК-1	принципы сбора, отбора и обобщения информации	соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности	работы с информационными источниками, опыт научного поиска, создания научных текстов	Оценка результатов практического занятия.
	ОПК-4	новые научные принципы и методы исследований	применять на практике новые научные принципы и методы исследований	применения новых научных принципов и методов исследования для решения профессиональных задач	

Критерии и шкалы оценивания

1. Практическое занятие

Баллы	Критерии оценивания
11	<p>— Обучающийся выполнил все задания практического занятия в полном объеме. Работа полностью соответствует заданию, не имеет значительных недостатков.</p> <p>— Обучающийся дает четкие, полные, правильные ответы на дополнительные вопросы по технологиям, применяемым в процессе проведения практического занятия и техническим операциям, которые необходимо было выполнить для получения нужного результата.</p>
10	<p>— Обучающийся выполнил все задания практического занятия в полном объеме. Работа соответствует заданию, но имеет значительные недостатки.</p> <p>— Обучающийся затрудняется дать полностью четкие, полные, правильные ответы на дополнительные вопросы по технологиям, применяемым в процессе проведения практического занятия и техническим операциям, которые необходимо было выполнить для получения нужного результата.</p>
9	<p>— Обучающийся не выполнил все задания практического занятия в полном объеме.</p> <p>— Обучающийся не может дать четкие, полные, правильные ответы на дополнительные вопросы по технологиям, применяемым в процессе проведения практического занятия и техническим операциям, которые необходимо было выполнить для получения нужного результата.</p>

Темы практических занятий, описание структуры и содержания, вопросы для подготовки к практическому занятию представлены в методических указаниях к практическим занятиям по дисциплине.

2. Контрольная работа

Баллы	Критерии оценивания
7	<p>— Обучающийся выполнил задание в полном объеме. Работа полностью соответствует заданию, не имеет значительных недостатков.</p> <p>— Обучающийся дает четкие, полные, правильные ответы на дополнительные вопросы по технологиям, применяемым в процессе проведения практического занятия и техническим операциям, которые необходимо было выполнить для получения нужного результата.</p>
4	<p>— Обучающийся выполнил задание в полном объеме. Работа соответствует заданию, но имеет значительные недостатки.</p> <p>— Обучающийся затрудняется дать полностью четкие, полные, правильные ответы на дополнительные вопросы по технологиям, применяемым в процессе проведения практического занятия и техническим операциям, которые необходимо было выполнить для получения нужного результата.</p>
1	<p>— Обучающийся не выполнил задание в полном объеме.</p> <p>— Обучающийся не может дать четкие, полные, правильные ответы на дополнительные вопросы по технологиям, применяемым в процессе проведения практического занятия и техническим операциям, которые необходимо было выполнить для получения нужного результата.</p>

В качестве контрольной работы обучающемуся выдаётся задание на подготовку научно-исследовательской (обзорной) работы.

Вопросы к контрольной работе по дисциплине «Научно-исследовательский семинар (Методология научных исследований и разработок)»

Сущность методологии исследования

Принципы и проблема исследования

Разработка гипотезы и концепции исследования

Процессуально-методологические схемы исследования

Научные методы познания в исследованиях

Методы научных исследований. Общенаучные методы.

Методы научных исследований. Специальные методы.

Библиография. Форматы оформления ссылок. Библиотечные менеджеры.

Виды научных исследований. Разница между поисковыми, фундаментальными и прикладными исследованиями.

История и характерные черты научно-технических революций.

Уровень научно-технического развития стран мира (системы показателей, сравнительная оценка).

Проблема добросовестности в научных исследованиях.

Основные виды научных публикаций (статьи, тезисы, отчеты о НИР, монографии и др.)

Технологические уклады. Общая характеристика. Причины перехода. Современное состояние в ведущих странах мира.

Наиболее значимые научные открытия последних (5) лет.

История науки и техники.

Развитие научно-исследовательской работы студентов (НИРС) в России.

Оценка результатов научной деятельности (количественная и качественная) (наукометрия/квалиметрия)

Характерные черты и структурные элементы современной научной статьи (IMRAD).

Другая тема по согласованию с преподавателем

Пример контрольной работы по дисциплине «Научно-исследовательский семинар (Методология научных исследований и разработок)»

ВВЕДЕНИЕ

Инструменты web scraping (парсинг) разработаны для извлечения, сбора любой открытой информации с веб-сайтов. Эти ресурсы нужны тогда, когда необходимо быстро получить и сохранить в структурированном виде любые данные из интернета. Парсинг сайтов – это новый метод ввода данных, который не требует повторного ввода или копипастинга.

Такого рода программное обеспечение ищет информацию под контролем пользователя или автоматически, выбирая новые или обновленные данные и сохраняя их в таком виде, чтобы у пользователя был к ним быстрый доступ. Например, используя парсинг можно собрать информацию о продуктах и их стоимости на сайте Amazon.

Ниже рассмотрим варианты использования веб-инструментов извлечения данных и десятку лучших сервисов, которые помогут собрать информацию, без необходимости написания специальных программных кодов. Инструменты парсинга могут применяться с разными целями и в различных сценариях, рассмотрим наиболее распространенные случаи использования, которые могут вам пригодиться. И дадим правовую оценку парсинга в России.

Каким бы методом социологические данные добыты ни были, сами по себе они ещё не позволяют сделать обобщённые выводы, выявить тенденции, проверить гипотезы - одним

словом, решить задачи, поставленные в программе исследования. Для того чтобы полученная первичная информация начала давать реальные результаты, её необходимо путём обработки превратить в соответствующий вид, а далее обобщить, проанализировать и научно проинтерпретировать.

Сервисы для электронного сбора данных разработаны специально для извлечения информации с веб-сайтов. Эти инструменты полезны для тех, кто пытается собрать различные данные из интернета. Это новый метод записи данных, который не требует повторного набора текста или копирования.

1. Правомерность применения программных средств сбора данных онлайн

Парсинг – это автоматизированный процесс извлечения данных с чужого веб-сайта. Но стоит разобраться, действительно ли это один из самых полезных инструментов ИТ для сбора данных или ловушка, влекущая неизбежные проблемы с законом? Парсинг мог бы непременно стать одним из совершеннейших способов добычи контента по всей сети, но к нему прилагается оговорка: с этим инструментом очень сложно разобраться с юридической стороны.

Парсинг – это процесс, посредством которого автоматизированная часть программного обеспечения извлекает данные веб-сайта, «прочесывая» многочисленные страницы. Поискковые системы как Google и Bing делают нечто подобное, когда индексируют веб-страницы, а парсинговые механизмы идут дальше и преобразовывают информацию в формат, который позволяет этими данными пользоваться, заносить в базы или электронные таблицы.

Парсинг – не то же самое, что API. Например, компания может открыть доступ к API, чтобы позволить другим системам взаимодействовать с ее данными; при этом качество и количество доступных данных через API, как правило, ниже, чем можно получить, используя парсинг. Кроме того, парсинг предоставляет более актуальную информацию, чем через API, и значительно проще настраивается со структурной точки зрения.

Области применения «парсинговой» информации очень многочисленны. Спортивный журналист может использовать парсинг, чтобы исследовать бейсбольную статистику для статьи. Или, например, в электронной коммерции можно извлекать наименования товаров и цены на них из разных источников для последующего анализа (как пример в России — открытый сервис парсинга и мониторинга цен конкурентов xmldatafeed.com). Вправе ли организация осуществлять автоматизированный сбор информации, размещенной в открытом доступе на сайтах в сети интернете (парсинг)?

В соответствии с действующим в Российской Федерации законодательством разрешено всё, что не запрещено законодательством. Парсинг является законным, в том случае, если при его осуществлении не происходит нарушений установленных законодательством запретов. Таким образом, при автоматизированном сборе информации необходимо соблюдать действующее законодательство.

Законодательством Российской Федерации установлены следующие ограничения, имеющие отношение к сети интернет:

1. Не допускается нарушение Авторских и смежных прав.
2. Не допускается неправомерный доступ к охраняемой законом компьютерной информации.
3. Не допускается сбор сведений, составляющих коммерческую тайну, незаконным способом.

4. Не допускается заведомо недобросовестное осуществление гражданских прав (злоупотребление правом).
5. Не допускается использование гражданских прав в целях ограничения конкуренции.

Из вышеуказанных запретов следует, что организация вправе осуществлять автоматизированный сбор информации, размещенной в открытом доступе на сайтах в сети интернет если соблюдаются следующие условия:

1. Информация находится в открытом доступе и не защищается законодательством об авторских и смежных правах.
2. Автоматизированный сбор осуществляется законными способами.
3. Автоматизированный сбор информации не приводит к нарушению в работе сайтов в сети интернет.
4. Автоматизированный сбор информации не приводит к ограничению конкуренции.

При соблюдении установленных ограничений Парсинг является законным, но стоит упомянуть громкие дела:

2009: Facebook. В 2009 году Facebook подал в суд на Power.com – сайт, который объединял различные социальные сети в один централизованный ресурс, – когда последний включил Facebook в свой сервис. Поскольку Power.com парсил контент Facebook, вместо того чтобы придерживаться установленных стандартов гиганта, Facebook предъявил иск на основании нарушения авторского права. Компания Facebook обвиняла Power.com в копировании веб-сайта Facebook в процессе извлечения информации о пользователях. Facebook утверждала, что этот процесс является прямым и косвенным нарушением авторского права. Решение суда было в пользу Facebook, и с этого времени решения относительно законности парсинга начали приниматься в пользу авторов содержимого сайтов.

Даже если парсер игнорирует контрафактный контент в процессе поиска общедоступной информации, его действия могут характеризоваться как нарушение авторского права, потому что технически контрафактный контент все равно «копируется».

2011-2014: Ауэрнхаймер. В 2010 году хакер Эндрю Ауэрнхаймер нашел брешь в системе безопасности на веб-сайте AT&T и извлек адреса электронной почты пользователей, которые посещали сайт со своих iPad. Воспользовавшись недостатком системы безопасности и парсингом, Ауэрнхаймер смог получить доступ к тысячам электронных адресов с сайта AT&T. Ауэрнхаймер был признан виновным в несанкционированном доступе к серверу AT&T и присвоении чужих данных.

Использование парсинга для извлечения конфиденциальной личной информации может привести к обвинению, даже если эта информация была номинально общедоступной. Можно попробовать убедить суд, что ни пароли, ни коды не взламывались для получения доступа к информации, тем не менее, это опасная территория.

2013: Meltwater. Meltwater – компания-разработчик программного обеспечения, чей продукт Global Media Monitoring, использует парсинг для сбора новостей. Ассошиэйтед Пресс подала в суд на Meltwater за парсинг статей, некоторые из которых были защищены авторским правом, и за незаконное присвоение новостей. Факты нельзя защищать авторским правом, но суд решил, что сами статьи и авторское изложение фактов копировать незаконно. Кроме того, использование статей компанией Meltwater не соответствовало установленным стандартам. Авторский контент не всегда можно парсить!

2014: QVC. В 2014 году QVC (известный телевизионный ритейлер) и Resultly (приложение-магазин) судились из-за того, что QVC назвал «чрезмерным парсингом». Обвинение QVC было в том, что Resultly маскировала своих поисковых роботов, чтобы скрыть исходный IP-адрес, таким образом QVC не могли блокировать нежелательных для них парсеров. Из-за

того, что боты были достаточно агрессивны к серверам QVC, произошла перегрузка с отключением электричества, которое повлекло ущерб в 2 миллиона долларов. Суд оправдал Resultly, постановив, что намерений нанести ущерб не было.

Есть рекомендации, которых стоит придерживаться, если используется парсинг: Извлекаемый контент не должен быть защищен авторским правом:

- Процесс парсинга не должен мешать работе сайта, который подвергается парсингу
- Парсинг не должен нарушать условия использования сайта
- Парсер не должен извлекать личную (персональную) информацию пользователя
- Контент, который подвергается парсингу, должен отвечать стандартам правомерного использования

2. Основные случаи использования программных средств сбора данных онлайн

Инструменты веб-скрейпинга могут использоваться для неограниченных целей в различных сценариях, но мы расскажем об общих случаях использования, применимых к обычным пользователям.

Функции сервисов для электронного сбора данных;

- структурирования найденной информации и представление её в удобном формате;
- сбор данных для исследования рынка;
- извлечение контактной информации;
- поиск работы или сотрудников;
- отслеживание цен с рынков.

1. Сбор данных для исследования рынка

Инструмент для сбора данных может помочь вам быть в курсе того, куда ваша компания или отрасль будет двигаться в ближайшие шесть месяцев, служа мощным инструментом для исследования рынка. Инструменты позволяют получать данные от нескольких поставщиков аналитической информации и фирм, занимающихся исследованием рынка, и объединять их в одном месте для удобства использования и анализа.

2. Извлечение контактной информации

Эти инструменты также могут использоваться для извлечения данных, таких как электронные адреса и номера телефонов с различных веб-сайтов, что позволяет получить список поставщиков, производителей и других лиц, представляющих интерес для вашего бизнеса или компании, наряду с их соответствующими контактными данными.

3. Загрузка решений из Stack Overflow

Используя инструмент веб-скрейпинга, можно также загружать решения для автономного чтения или хранения, собирая данные с нескольких сайтов (включая Stack Overflow и другие веб-сайты Q&A). Это снижает зависимость от активных подключений к интернету, поскольку ресурсы доступны, несмотря на наличие сети.

4. Поиск работы или сотрудников

Для компаний, которые активно ищут новых кандидатов в свою команду, или для соискателей, которые ищут определенную работу или вакансию, эти инструменты также отлично работают, помогая легко получать данные на основе различных применяемых фильтров и делать это эффективно без необходимости ручного поиска.

5. Отслеживание цен с нескольких рынков

Если вы занимаетесь онлайн-покупками и любите активно отслеживать цены на товары, которые вы ищете на разных рынках и в онлайн-магазинах, то вам определенно нужен инструмент для веб-скрейпинга в Интернете.

В будущем информационную роль наверняка ждет большое распространение. Конечно, она и сейчас распространена, но не так как хотелось бы. Например, было бы неплохо в наше время практиковать сбор информации на крупных предприятиях, как это делается за рубежом.

Это сильно помогло бы принимать правильные решения руководителям, которые толком не знают о том, что творится в коллективах их предприятий. Зачастую принимаются решения, которые противоречат мнению коллективов и общественности, чего помогли бы избежать сбор нужной информации.

3. Лучшие современные программные средства сбора данных онлайн

Давайте взглянем на 10 лучших инструментов для веб-скрейпинга. Некоторые из них бесплатны, а другие имеют пробные периоды и премиальные планы. Изучите детали, прежде чем подписаться на один из них для ваших потребностей.

Import.io

Import.io предлагает конструктор для формирования ваших собственных наборов данных, просто импортируя данные с определенной веб-страницы и экспортируя их в CSV. Вы можете легко собрать данные с тысячи веб-страниц за считанные минуты без единой строчки кода и построить более 1000 API-интерфейсов на основе ваших требований.

Инструмент использует передовые технологии для получения миллионов данных каждый день, которые предприятия могут использовать за небольшую плату. Наряду с веб-инструментом, он также предлагает бесплатные приложения для Windows, Mac OS X и Linux, которые позволяют собирать и извлекать данные, загружать и синхронизировать их с онлайн-аккаунтом.

	newimgsmall...	productpaddl...	productpaddl...	proding_lma...	prodname_va...	proddesc_val...	prodprice_pri...	unitprice_label	proddimensio...	proddimensio...	proddimensio...
1	STUVA Storage...	79.99		STUVA	Storage bench	79.99	Unit price	Width: 35 3/8 "	35	Depth: 19 5/8 "	
2	POÄNG Chair \$...	159		POÄNG	Chair	159	Unit price	Depth: 32 1/4 "	32	Seat width: 22 "	
3	TORBJÖRN Swi...	39.99		TORBJÖRN	Swivel chair	39.99	Unit price	Tested for: 242...	242	Depth: 23 5/8 "	
4	VOLMAR Swive...	209		VOLMAR	Swivel chair wit...	209	Unit price	Tested for: 242...	242	Seat width: 16 ...	
5	New EKTORP Chair c...	99		EKTORP	Chair cover	99	Unit price				
6	POÄNG Rockin...	229		POÄNG	Rocking chair	229	Unit price	Width: 26 3/4 "	26	Depth: 37 "	
7	SKRUVSTA Swi...	149		SKRUVSTA	Swivel chair	149	Unit price	Tested for: 242...	242	Min. height: 31 ...	
8	FLINTAN / NOM...	89.99		FLINTAN / NOM...	Swivel chair wit...	89.99	Unit price	Tested for: 242...	242	Width: 29 1/8 "	

Webhose.io

Webhose.io предоставляет прямой доступ к структурированным данным в режиме реального времени, сканируя тысячи онлайн-источников. Инструмент поддерживает извлечение веб-данных на более чем 240 языках и сохранение выходных данных в различных форматах, включая XML, JSON и RSS.

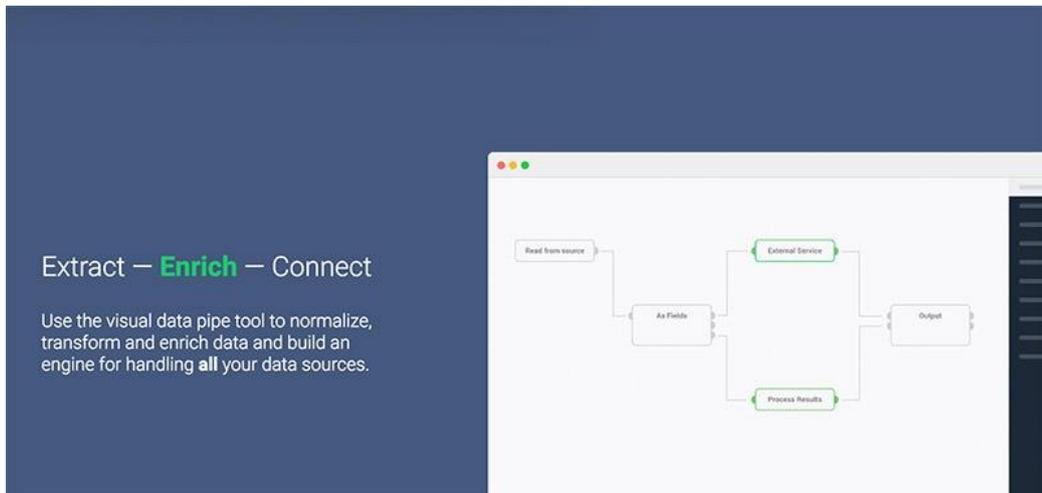
Webhose.io — это веб-приложение на основе браузера, использующее эксклюзивную технологию сканирования данных для сбора огромных объемов данных из нескольких каналов в одном API. Он предлагает бесплатный план для 1000 запросов в месяц и премиальный план с ежемесячной платой 50 долларов на 5000 запросов в месяц.



Dexi.io (ранее известный как CloudScrape)

CloudScrape поддерживает сбор данных с любого веб-сайта и не требует загрузки, как Webhose. Он предоставляет браузерный редактор для настройки сканеров и извлечения данных в режиме реального времени. Вы можете сохранить собранные данные на облачных платформах, таких как Google Drive, или экспортировать их в формате CSV или JSON.

CloudScrape также поддерживает анонимный доступ к данным, предлагая набор прокси-серверов для скрытия вашей личности. Инструмент хранит ваши данные на своих серверах в течение двух недель перед их архивированием. Данный инструмент для веб-скрейпинга предлагает 20 бесплатных часов для сбора данных, после чего будет стоить 29 долларов в месяц.



Scrapinghub

Scrapinghub — это облачный инструмент для извлечения данных, который помогает тысячам разработчиков получать ценные данные. Scrapinghub использует Crawlera, интеллектуальный прокси-ротатор, который поддерживает сканирование противодействий ботов для легкого сканирования огромных сайтов или страниц, защищенных ботами.

Scrapinghub преобразует всю веб-страницу в организованный контент. Группа экспертов всегда готова помочь, если ее конструктор сканирования не сможет удовлетворить ваши требования. Базовый бесплатный план дает вам доступ к одному одновременному сканированию, а премиум-план за 25 долларов в месяц обеспечивает доступ к четырем параллельным сканированиям.

Job	Spider	Version	Pages	Items	Errors	Log	Wait Time	Added
87/5945	eWatches	auto					3 min, 12 sec	2013-08-07
32/5979	WeedForWine	auto					3 min, 12 sec	2013-08-07 16:0
73/5965	WengerWatchOutD	auto					3 min, 12 sec	2013-08-07 16:0
72/5972	WaltonDeatohtheDay	auto					3 min, 13 sec	2013-08-07 16:0
75/5966	Vanns	auto					3 min, 13 sec	2013-08-07 16:0
53/6115	Tipp_Travel	auto					3 min, 14 sec	2013-08-07 16:0
76/5966	Titleded	auto					3 min, 14 sec	2013-08-07 16:02:04 UTC
64/5985	TigerDirect	auto					3 min, 14 sec	2013-08-07 16:02:04 UTC
83/5966	ThingFling	auto					3 min, 15 sec	2013-08-07 16:02:04 UTC
74/5967	TheWineSpies	auto					3 min, 15 sec	2013-08-07 16:02:03 UTC

Job	Spider	Version	Pages	Items	Errors	Log	Runtime	Started	Last Activity
89/5974	CowBoom	auto	0	0	0	0	5 sec	2013-08-07 16:05:13 UTC	less than a minute ago
119/5838	MacZOT	auto	0	0	0	0	10 sec	2013-08-07 16:05:09 UTC	less than a minute ago
24/5974	Cassafesta	auto	0	0	0	9	1 min, 5 sec	2013-08-07 16:04:14 UTC	less than a minute ago
110/5946	1SaleDay_Flash	auto	62	0	0	10	1 min, 54 sec	2013-08-07 16:03:25 UTC	less than a minute ago
102/541	RuKLaLa	auto	154	779	0	14	3 min, 30 sec	2013-08-07 16:01:49 UTC	less than a minute ago
103/274	OneKingLine	auto	405	0	0	25	4 min, 42 sec	2013-08-07 16:00:36 UTC	less than a minute ago

ParseHub

ParseHub создан для сканирования одного или нескольких веб-сайтов с поддержкой JavaScript, AJAX, сессий, файлов cookie и перенаправления. Приложение использует технологию машинного обучения для распознавания наиболее сложных документов в Интернете и создает выходной файл на основе требуемого формата данных.

ParseHub, кроме веб-приложения, также доступен в качестве бесплатного настольного приложения для Windows, Mac OS X и Linux, которое предлагает базовый бесплатный план,

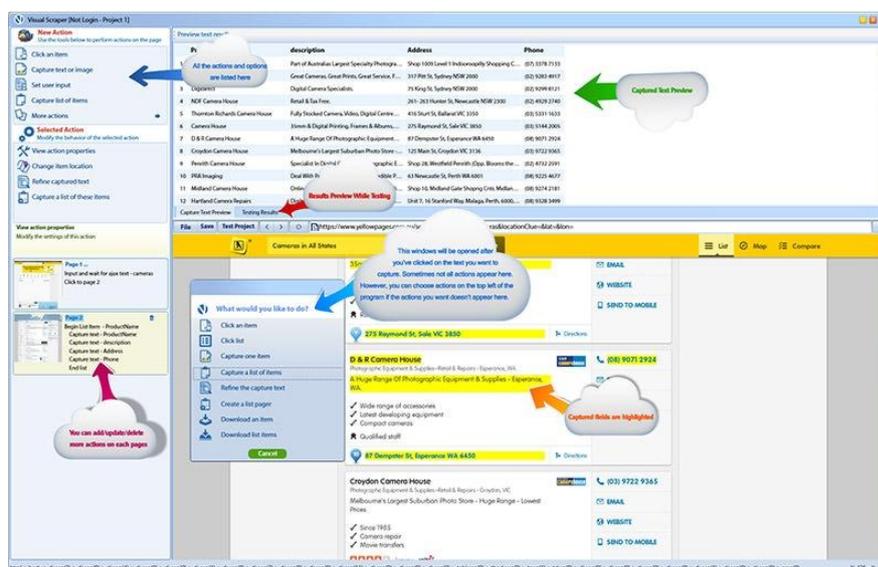
охватывающий пять проектов для сканирования. Этот сервис предлагает премиальный план за 89 долларов в месяц с поддержкой 20 проектов и 10 000 веб-страниц для сканирования.



VisualScaper

VisualScaper — еще одно программное обеспечение для извлечения веб-данных, которое можно использовать для сбора информации из Интернета. Программное обеспечение помогает вам извлекать данные с нескольких веб-страниц и получать результаты в режиме реального времени. Кроме того, вы можете экспортировать полученные данные в различные форматы, такие как CSV, XML, JSON и SQL.

Вы можете легко собирать и управлять веб-данными с помощью простого и удобного интерфейса. VisualScaper поставляется в бесплатных, а также премиальных планах, начиная с 49 долларов в месяц, с доступом к более 100 тысячам страниц. Бесплатное приложение, похожее на Parsehub, доступно для Windows с дополнительными пакетами C++.

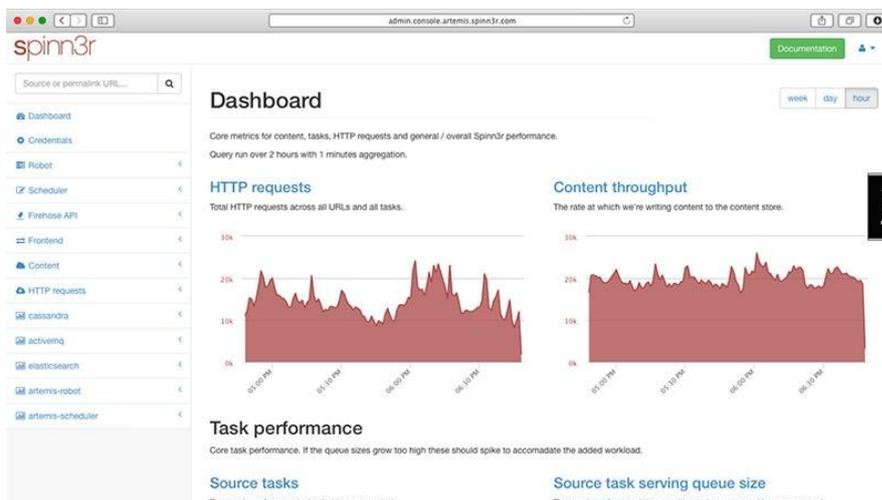


Spinn3r

Spinn3r позволяет получать целые данные из блогов, новостных сайтов и сайтов социальных сетей, а также каналов RSS и ATOM. Spinn3r поставляется с API Firehouse, который

управляет 95% работ по индексированию. Он предлагает расширенную защиту от спама, которая удаляет спам и неуместное использование языка, тем самым повышая безопасность данных.

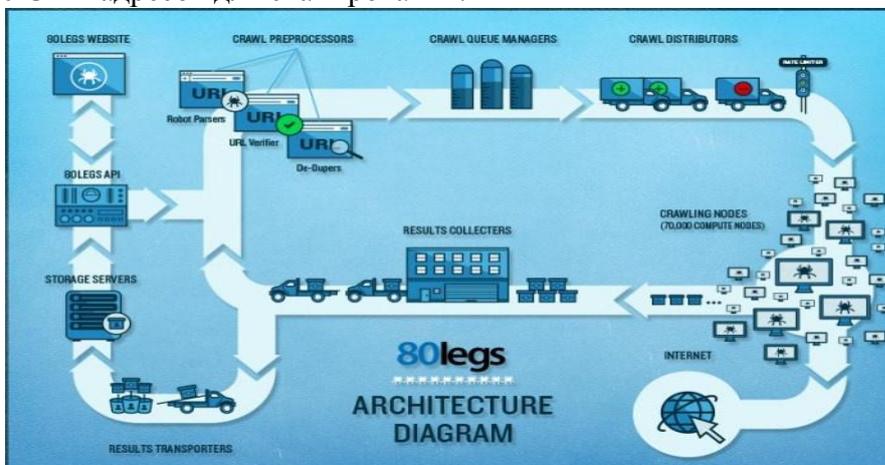
Spinn3r индексирует контент, аналогично Google, и сохраняет извлеченные данные в файлы JSON. Инструмент постоянно сканирует сеть и находит обновления из нескольких источников, чтобы получать публикации в реальном времени. Консоль администратора позволяет контролировать сканирование, а полнотекстовый поиск дает возможность выполнять сложные запросы к необработанным данным.



80legs

80legs — это мощный, но гибкий инструмент для сканирования веб-страниц в Интернете, который можно настроить в соответствии с вашими потребностями. Он поддерживает извлечение огромных объемов данных, а также возможность мгновенной загрузки извлеченных данных. Разработчики инструмента утверждают, что он сканирует более 600 000 доменов и используется такими крупными компаниями, как MailChimp и PayPal.

Его 'Datafiniti' позволяет быстро искать все данные. 80legs обеспечивает высокопроизводительное сканирование в Интернете, которое работает быстро и извлекает необходимые данные за считанные секунды. Вам предлагается бесплатный план на 10 000 URL-адресов для сканирования, который может быть обновлен до премиального плана за 29 долларов в месяц на 100 000 URL-адресов для сканирования.



Scraper

Scraper — это расширение для Chrome с ограниченными функциями извлечения данных, но оно полезно для онлайн-исследований и экспорта данных в электронные таблицы Google. Этот инструмент предназначен для начинающих, а также экспертов, которые смогут легко копировать данные в буфер обмена или сохранять в электронных таблицах с помощью OAuth.

Scraper является бесплатным инструментом, который работает прямо в вашем браузере и автоматически генерирует меньшие XPath для определения URL для сканирования. Он не предлагает простоты автоматического сканирования или сканирования ботов, как Import, Webhose и другие, но он довольно полезен для новичков, так как вам не придется разбираться с запутанной конфигурацией.

ЗАКЛЮЧЕНИЕ

Программное обеспечение для сбора данных ищет новые данные вручную или автоматически, извлекая новые или обновлённые данные, а также сохраняя их для легкого доступа. Например, с помощью такого сервиса можно собирать информацию о продуктах и их ценах. Данные можно преобразовать в таблицы, файлы, базы данных.

Но, хоть парсинг и, несомненно, мощный инструмент, когда дело касается правовых вопросов, могут появиться трудности. Поскольку в процессе парсинга изначально существующий контент из разных источников присваивается тем, кто этот инструмент использует, появляются этические и юридические затруднения.

На сегодняшний день в среде парсинга нет четко очерченных юридических рамок, это состояние постоянного движения, но можно попытаться приблизительно обрисовать зоны наибольшего риска. Ниже в общих чертах описываются самые яркие случаи судебных разбирательств, имевшие место в США, и ставшие прецедентными.

Электронный сбор данных имеет плюсы и минусы.

Плюсы:

- легкость сканирования огромных сайтов или страниц;
- конвертация полученных данных в различные форматы;
- безопасность данных.

Минусы:

- довольно высокая стоимость.

В соответствии с действующим в Российской Федерации законодательством разрешено всё, что не запрещено законодательством. Парсинг сайтов является законным, в том случае, если при его осуществлении не происходит нарушений установленных законодательством запретов. Таким образом, при автоматизированном сборе информации необходимо соблюдать действующее законодательство.

Вопросы к экзамену по дисциплине

«Научно-исследовательский семинар (Методология научных исследований и разработок)»

1. Основы научно-исследовательской деятельности.
2. Наука как вид человеческой деятельности.
3. Сущность и структура науки как особого вида знания.
4. Типология научных исследований.

5. История и тенденции развития науки.
6. История науки.
7. Понятие научной революции.
8. Тенденции развития науки.
9. Законодательные основы научных исследований.
10. Законодательные акты, регламентирующие управление научной деятельностью.
11. Нормативные документы, регламентирующие организацию фундаментальных и прикладных исследований.
12. Акты правовой охраны интеллектуальной собственности ученых.
13. Правовая база выполнения квалификационных исследований.
14. Методологические основы научных исследований.
15. Понятие о методе и методологии исследования.
16. Уровни методологии.
17. Универсалии науки.
18. Типология методов научных исследований.
19. Логика научного исследования.
20. Алгоритм научного исследования.
21. Общий алгоритм проведения научного исследования.
22. Выбор направления и темы научного исследования.
23. Постановка научно-практической задачи (проблемы).
24. Разработка научной гипотезы.
25. Теоретические исследования.
26. Сущность теоретических исследований.
27. Методы проведения теоретических исследований.
28. Основы системного анализа.
29. Эмпирические исследования.
30. Сущность и виды эмпирических исследований.
31. Методы проведения эмпирических исследований.
32. Основы моделирования.
33. Основы теории эксперимента.
34. Сущность и виды эксперимента.
35. Основы теории эксперимента.
36. Планирование эксперимента.
37. Особенности проведения социального эксперимента.
38. Планирование и организация научных исследований.
39. Основы планирования научных исследований.
40. Перспективное и текущее планирование.
41. Организация фундаментальных научных исследований.
42. Организация научных исследований и конструкторской подготовки производства.
43. Аспирантская и магистерская подготовка.
44. Организация аспирантской подготовки.
45. Общие требования и организация магистерской подготовки.
46. Подготовка и оформление кандидатской и магистерской диссертаций.
47. Подготовка диссертации на соискание ученой степени кандидата наук.
48. Особенности подготовки магистерской диссертации.
49. Планирование диссертационного исследования.
50. Оформление диссертационной работы.
51. Публикация, внедрение и защита результатов диссертационного исследования.
52. Публикация результатов диссертационного исследования.
53. Внедрение результатов диссертационного исследования.
54. Обсуждение результатов научных исследований.
55. Защита диссертации.