ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**«САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО»**

Институт компьютерных наук и технологий

Высшая школа интеллектуальных систем и суперкомпьютерных технологий

Дисциплина «Цифровые ресурсы в научных исследованиях»

### ОТЧЕТ

по лабораторной работе №1

на тему

«Первичный полнотекстовый поиск источников информации в глобальной сети интернет»

Выполнил:

студент группы 3540901/02001

Бараев Дамир Рашидович

«\_\_» \_\_\_\_\_\_\_\_\_\_ 2020г., \_\_\_\_\_\_\_\_\_\_

(подпись)

(подпись)

Проверила:

Бендерская Елена Николаевна

«\_\_» \_\_\_\_\_\_\_\_\_\_ 2020г., \_\_\_\_\_\_\_\_\_\_

Санкт-Петербург 2020

**СОДЕРЖАНИЕ**

[1 Постановка задачи 3](#_Toc52742144)

[2.1. Задача 1 4](#_Toc52742145)

[2.2. Задача 2. 10](#_Toc52742146)

[2.3. Задача 3. 11](#_Toc52742147)

[2.4. Задача 4. 11](#_Toc52742148)

[2.5. Задача 5. 15](#_Toc52742149)

[2.6. Задача 6. 17](#_Toc52742150)

[2.7. Задача 7 19](#_Toc52742151)

[3. Заключение 24](#_Toc52742152)

# 1 Постановка задачи

Основной целью данной работы является полное освоение первичного полнотекстового поиска источников информации в сети интернет. Задачами лабораторной работы являются:

1. Формулировка основных ключевых слов и словосочетаний по теме аналитического отчета и выполнение поиска по ним, с дальнейшей оценкой его результатов.
2. Оценка качества и количества полученных при поиске по запросам ссылок и источников, уточнение и расширение запросов на основе оценки.
3. Сравнение ключевых слов, представленных в лучших найденных источниках.
4. Выполнение вышеупомянутых задач в другой поисковой системе.
5. Использование расширенного инструментария поисковых систем для поиска по составленным запросам.
6. Проведение работы с наиболее релевантным источником из найденных: переработка первичного текста во вторичный, составление аннотаций и информативного реферата текста.
7. Выполнение вышеупомянутых задач используя запросы на английском языке.

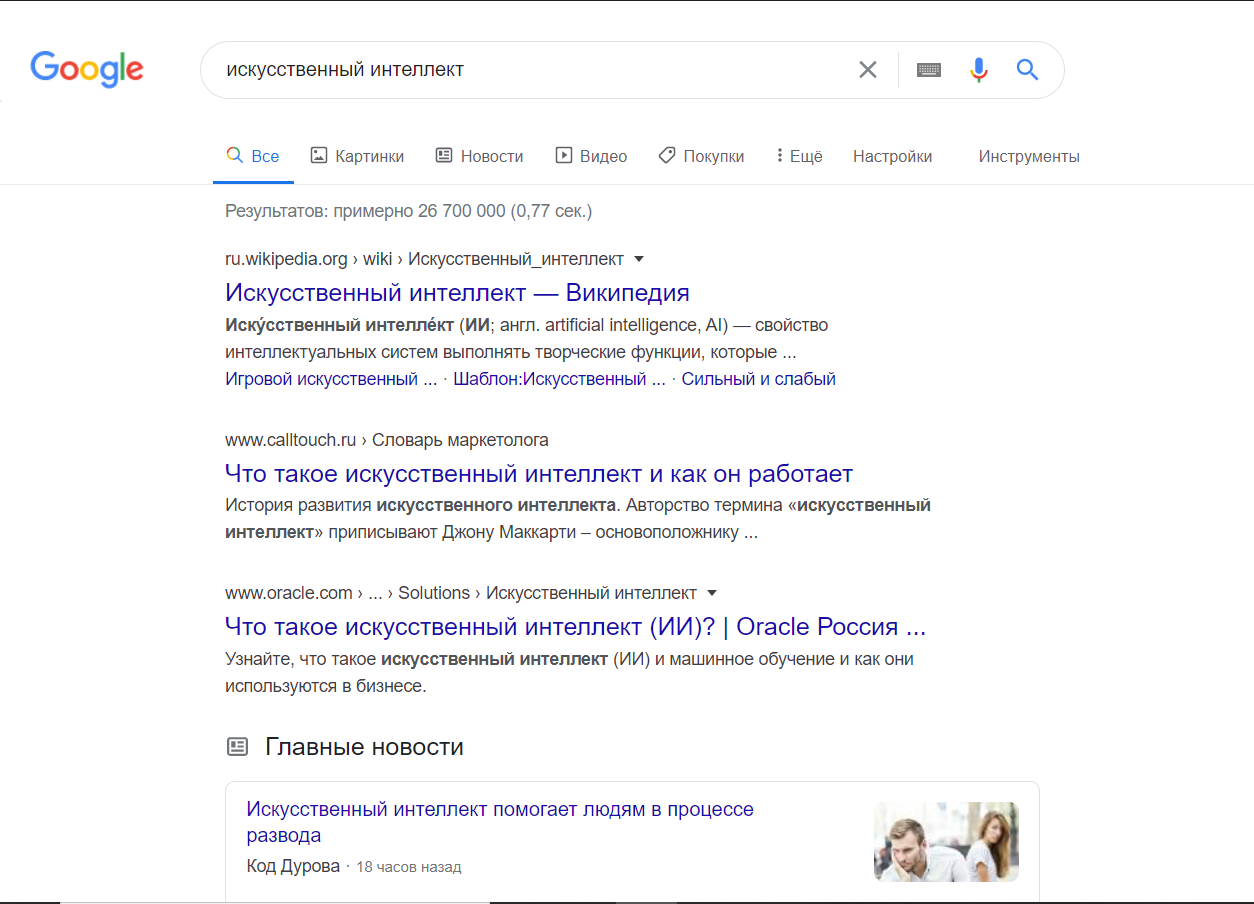
# 2.1. Задача 1

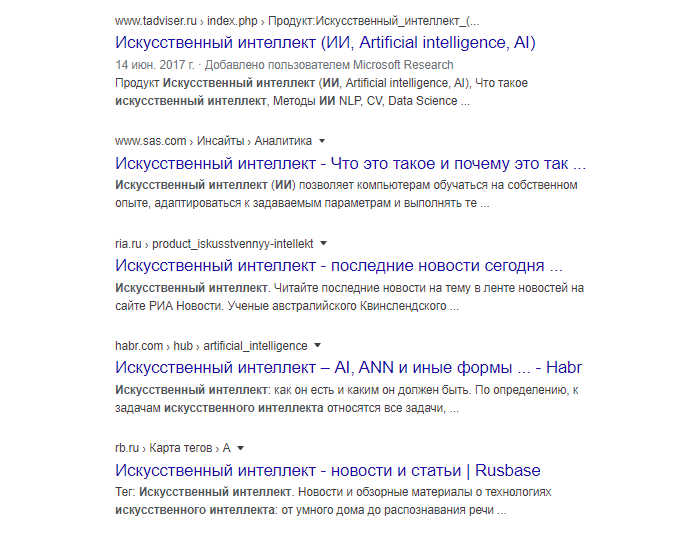
Тема аналитического отчета: «Программные среды разработки искусственного интеллекта».

Основные ключевые слова и словосочетания: искусственный интеллект, среды разработки ИИ, ПО искусственного интеллекта, программные среды разработки ИИ.

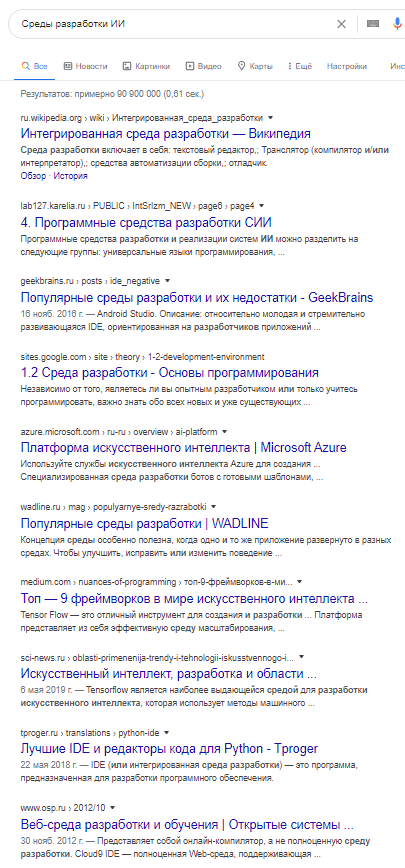
Оценки релевантности полученных ссылок и источников в поисковой системе **Google**:

1. Искусственный интеллект (Первичная оценка – 5 из 10)

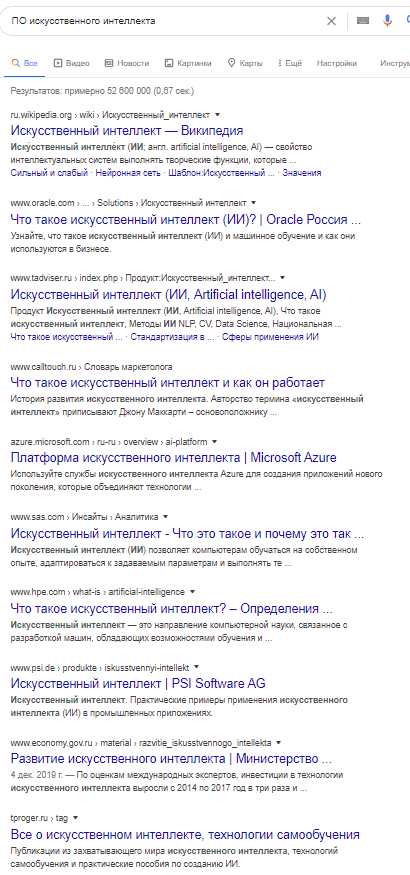




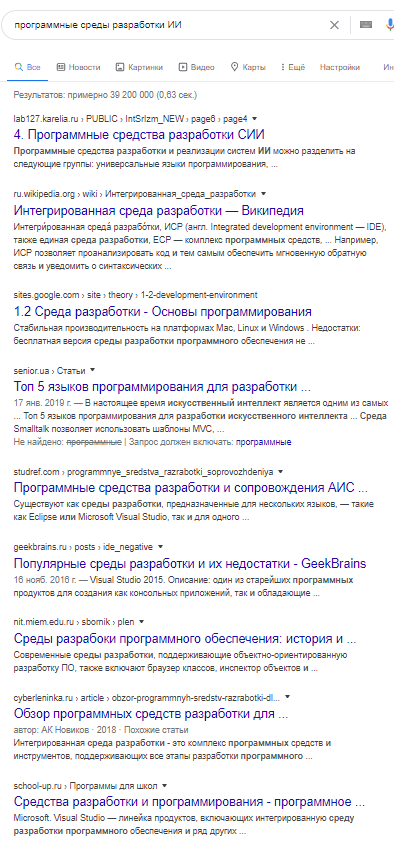
1. Среды разработки ИИ (Первичная оценка – 6 из 10)



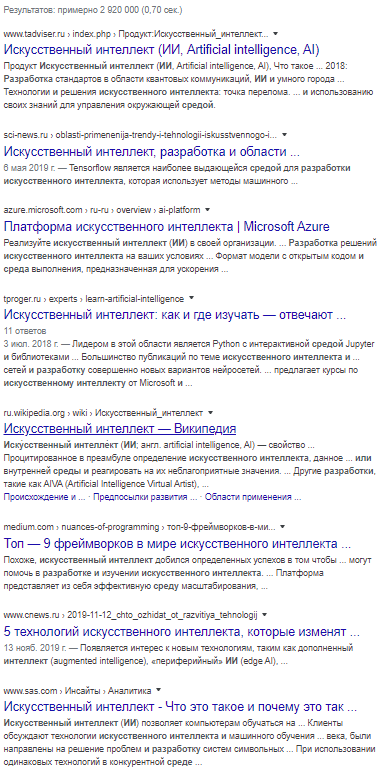
1. ПО искусственного интеллекта (4 из 10)



1. Поиск по теме аналитического отчета: программные среды разработки ИИ (6 из 10)



1. Поиск по всем ключевым словам (Первичная оценка – 4 из 10)



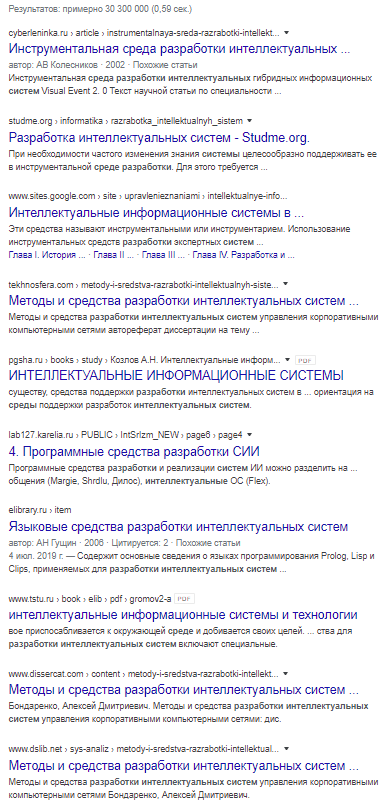
# 2.2. Задача 2.

Лучший вариант поискового запроса в предыдущем пункте – «Среды разработки ИИ». Оценка представленных источников:

Степень удовлетворенности качеством – 5 из 10. Дело в том, что в полной мере соответствуют теме только 3 источника и рассматривают они тему несколько поверхностно. Выходит, что ожиданиям они соответствуют, но не в должной мере, не говоря уже о количестве полученных источников.

Оценка достаточности представленных источников – 3 из 10, что соответствует количеству источников, с которыми можно работать для изучения темы.

Замена словосочетания «искусственный интеллект» на «интеллектуальная система» позволила найти конкретно новые источники:



Однако, оценка релевантности – 4 из 10, так как на представленном выше скриншоте – самые релевантные из них. При этом один из них является историческим.

# 2.3. Задача 3.

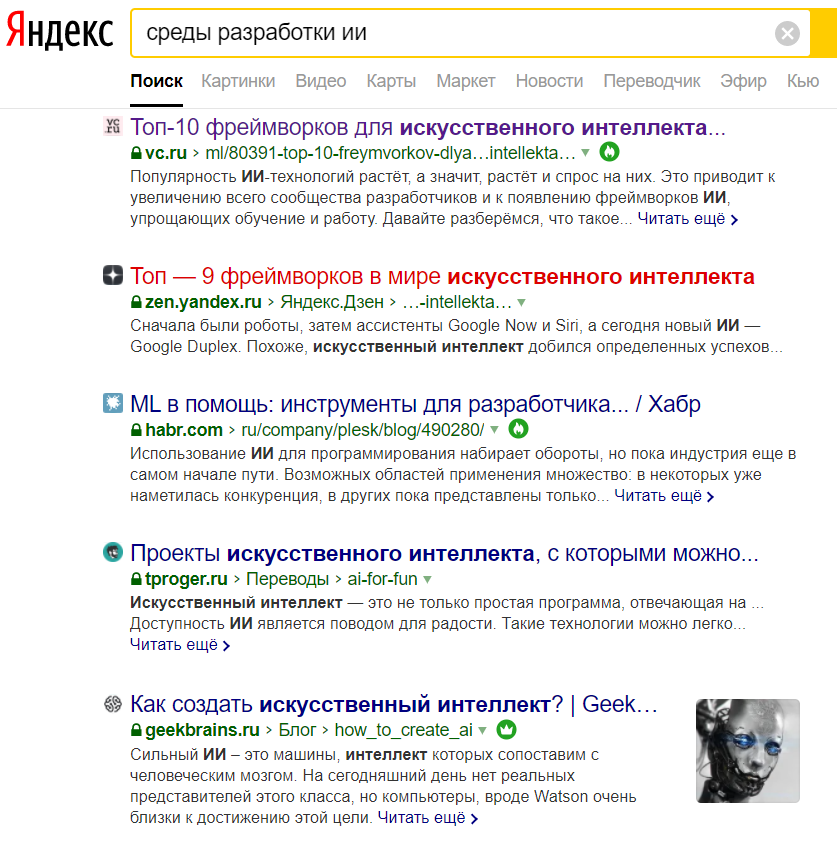
Лучшие найденные источники – *«Программные средства разработки СИИ»* (<http://lab127.karelia.ru/~ekostq/PUBLIC/IntSrIzm_NEW/page6/page4.html>), *«Программные средства для систем искусственного интеллекта» (*[*https://standartgost.ru/0/28323-programmnye\_sredstva\_dlya\_sistem\_iskusstvennogo\_intellekta*](https://standartgost.ru/0/28323-programmnye_sredstva_dlya_sistem_iskusstvennogo_intellekta)), *«Искусственный интеллект (ИИ, Artificial intelligence, AI)»* (<https://www.tadviser.ru/index.php/%D0%9F%D1%80%D0%BE%D0%B4%D1%83%D0%BA%D1%82:%D0%98%D1%81%D0%BA%D1%83%D1%81%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9_%D0%B8%D0%BD%D1%82%D0%B5%D0%BB%D0%BB%D0%B5%D0%BA%D1%82_(%D0%98%D0%98,_Artificial_intelligence,_AI)>)

Ключевые слова, которые в них использовались: *искусственный интеллект, язык программирования, программные средства, системы ИИ, программные продукты, технологии, средства создания систем ИИ, экспертные системы, варианты реализации искусственного интеллекта, введение в современные САПР.*

На основе сравнения этих ключевых слов и тех, что использовались мной, можно сделать вывод о необходимости коррекции и сокращения используемых ключевых слов, но при этом и необходимости формулировки общего поискового запроса, который будет нести в себе больше информации, которая выведет поиск к нужным результатам.

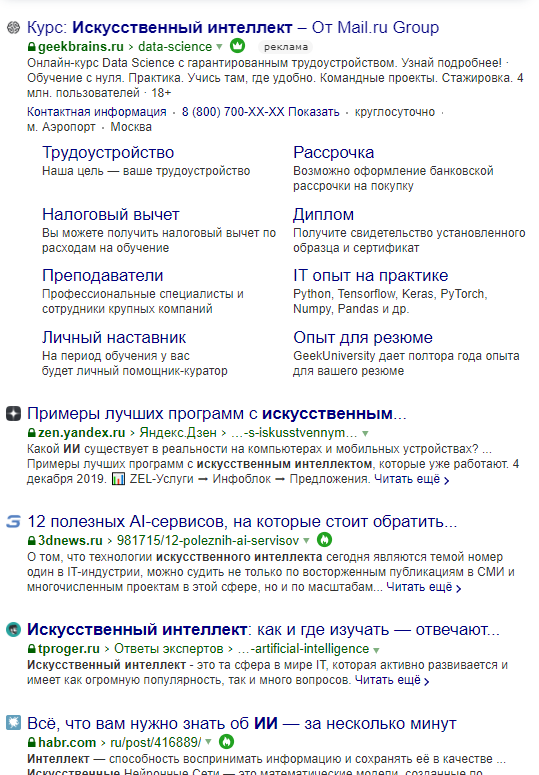
# 2.4. Задача 4.

В рамках данной задачи нужно повторить пункты 1-3, но при этом используя другую поисковую систему. Так, я буду использовать систему Yandex. По запросу «среды разработки ИИ» были получены в большинстве своём новые результаты, однако, остались и старые:



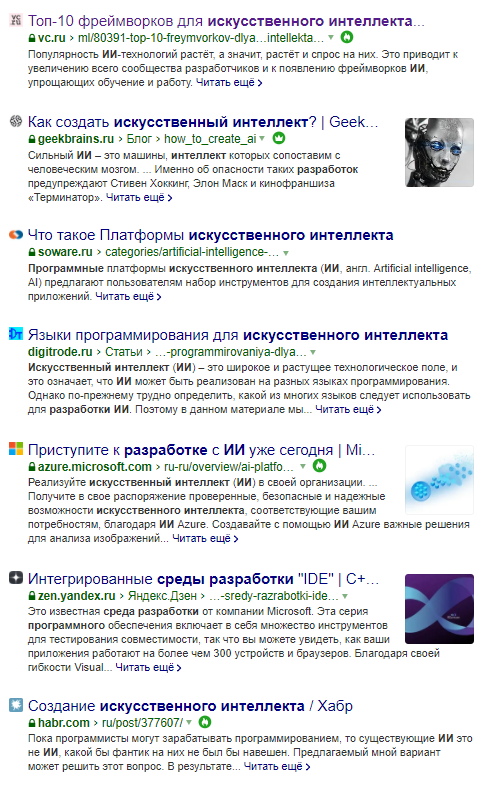
Хотя они и служат скорее примерами, раскрывающими тему, они, безусловно, могут быть полезными. Оценка – 5 из 10.

Запрос «ПО искусственного интеллекта» приходится оценивать на 3 из 10, потому что все полученные ссылки не особо соответствуют запросу, и вероятно, что они окажут небольшую помощь при составлении аналитического отчета.



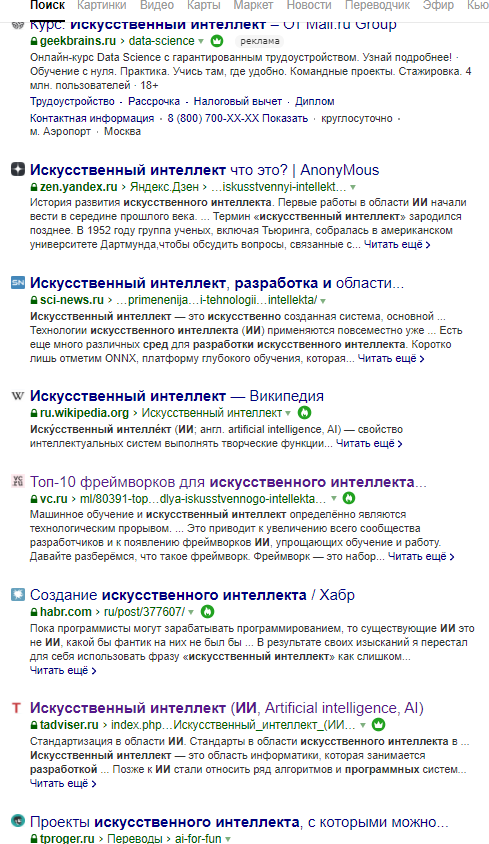
*Запрос – ПО искусственного интеллекта*

По запросу «Программные среды разработки ИИ» можно поставить оценку 5 из 10, так как несмотря на то, что многие результаты уже были получены ранее, появились и принципиально новые:



Запрос «Искусственный интеллект» не дал новых результатов.

Запрос по всем ключевым словам практически полностью отличается от предыдущего, но не все источники являются полезными для нашего отчета.

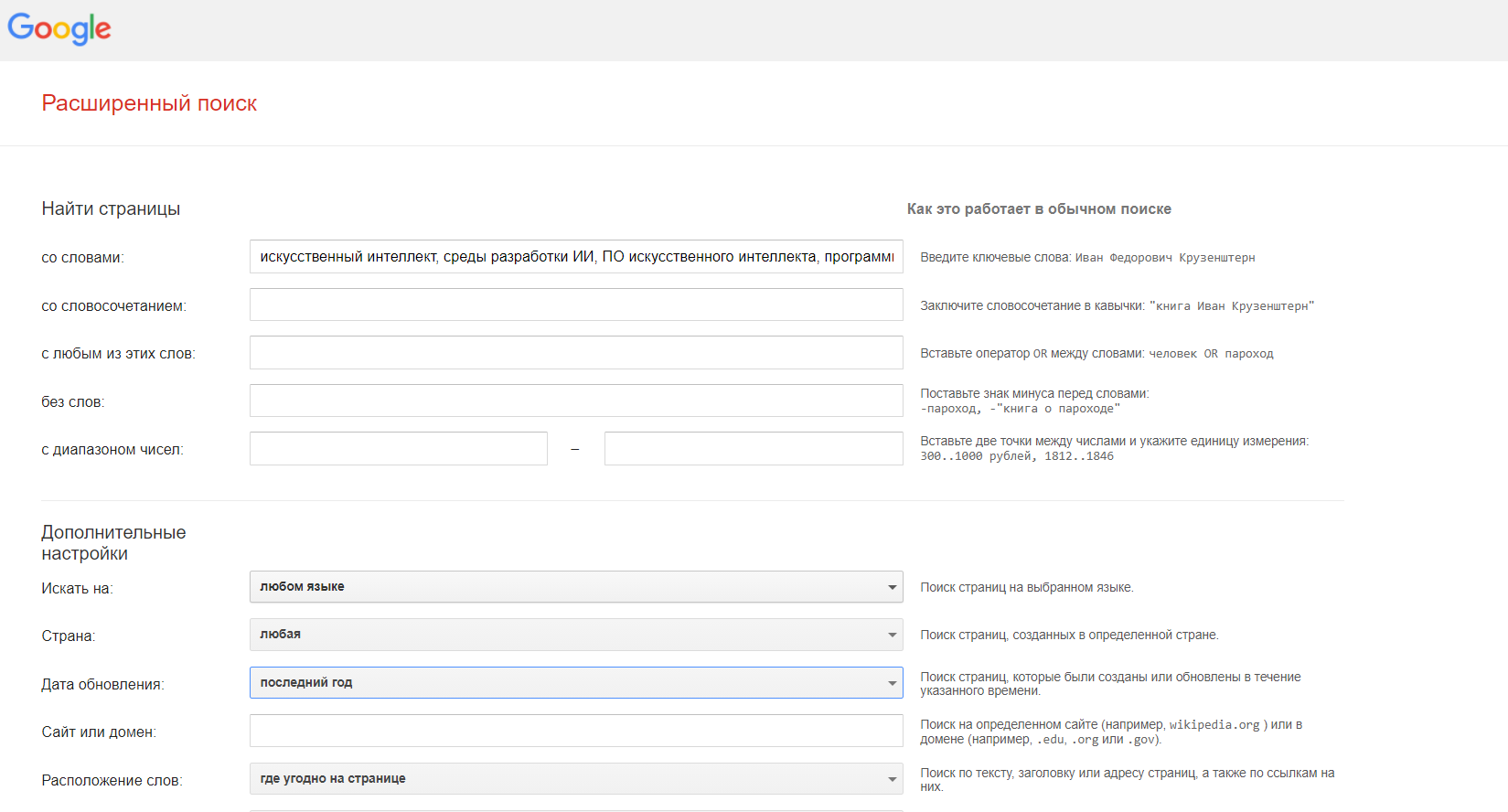


Запрос по теме аналитического отчета не принес новых результатов, более того, те, что оказались ранее не представленными, не несли в себе какой-то конкретной привязки к теме.

Таким образом, итоги по 2-3 пунктам также не меняются, потому что обе поисковые системы не имеют кардинальных отличий в представленных запросах. При этом стоит отметить, что в Yandex с первым же запросом были выданы лучшие результаты, которые нашлись в Google на третьем запросе.

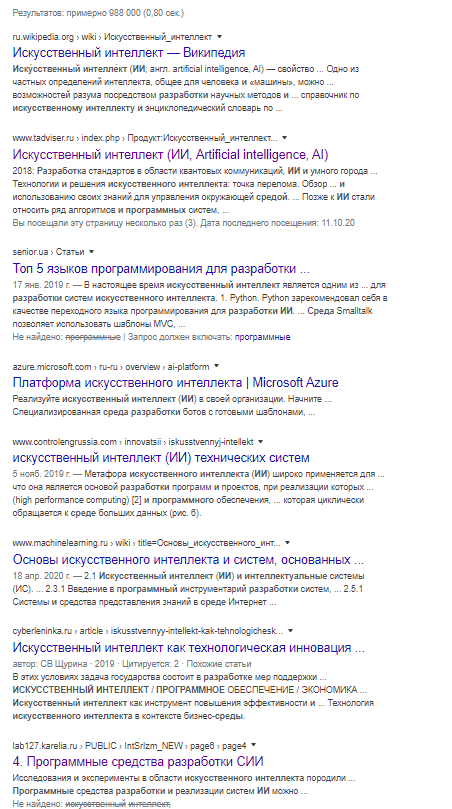
# 2.5. Задача 5.

Инструментарий расширенного поиска Google



Чтобы найти конкретные тексты, посвященные теме аналитического отчета, или с ключевыми словами, составленными по ней, в разделе дополнительных настроек я указал расположение слов «в тексте страницы». Язык и страна в данном контексте значения не имеют, т.к поиск ведется на русском языке. Дату обновления можно поставить за последний год, что позволит найти более актуальную информацию.

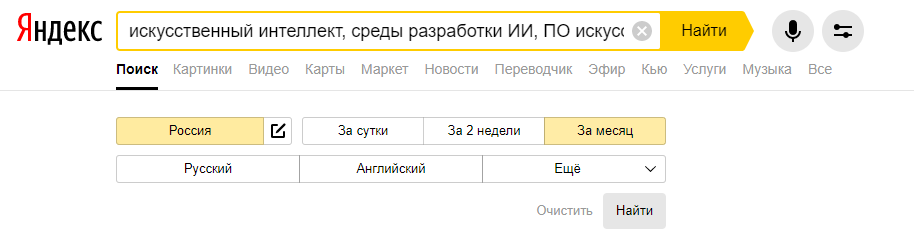
Результаты поиска за последний год не оказались продуктивными:



При этом изменение даты обновления на «за всё время» лучше результаты не сделало.

Так или иначе, использование расширенного поиска вновь меня привело к статье, которую я не раз находил в рамках предыдущих пунктов ([http://lab127.karelia.ru/~ekostq/PUBLIC/IntSrIzm\_NEW/page6/page4.html),](http://lab127.karelia.ru/~ekostq/PUBLIC/IntSrIzm_NEW/page6/page4.html) которую я и так внёс в список лучших результатов. Важно отметить, что это был единственный действительно полезный источник.

Интерфейс расширенного поиска в Яндекс несколько расстроил в сравнении с использованным ранее аналогом:



Особых результатов с ним достигнуто не было.

Таким образом, расширенный поиск обеих поисковых систем безусловно полезен в отдельных случаях, но, когда речь идет о поиске полезных и важных для составления аналитического отчета ресурсов, особой (и моментальной) помощи от него ожидать не стоит.

# 2.6. Задача 6.

Наиболее релевантными я посчитал учебник Т. А. Гаврилова и В. Ф. Хорошевский «Базы знаний интеллектуальных систем» и статью под редакцией Д. Н. Ушакова «На каком языке программирования разрабатывать искусственный интеллект?».

Особенностью учебника «Базы знаний интеллектуальных систем» является его практическая направленность. Освоив изложенный материал, студент или другой заинтересованный читатель сможет самостоятельно приступить к разработке интеллектуальной системы в роли инженера по знаниям. Акцент в учебнике сделан именно на работу со знаниями. Фактически он ориентирован на подготовку уникальных специалистов, спрос на которых на современном рынке высоких информационных технологий многократно превышает спрос на программистов.

Описательная аннотация: Необычность этого учебника связана также с подчеркнутой междисциплинарностью выбранного подхода, отказом от «клановости» отдельных научных школ и направлений. Этот учебник могут читать инженеры и математики, экономисты и биологи, программисты и медики. Он практически не требует предварительной подготовки в данной области знаний и рассчитан на широкий круг читателей, заинтересованных разработкой интеллектуальных систем, основанных на знаниях.

Мы отказались от излишней специализации в пользу широты изложения, нам

хотелось представить горизонты этой науки, а не прокопать в ней глубокий, но узкий туннель.

Аннотация-резюме: Учебник для технических вузов по входящим в различные дисциплины вопросам разработки интеллектуальных систем - развивающейся области информатики. Актуальность предмета определяется растущим применением инженерии знаний и системного анализа в различных областях деятельности.

Особенностью изложения является его практическая направленность: освоения имеющегося материала достаточно для начала самостоятельной работы над созданием интеллектуальной системы, основанной на знаниях.

В учебнике учтена все возрастающая роль Интернета, и потому подробно рассматривается применение инженерии знаний в Сети.

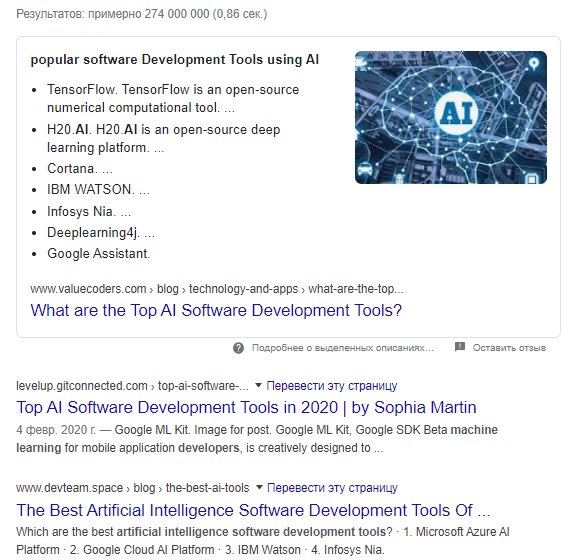
Ключевые слова: Введение в интеллектуальные системы. Разработка систем, основанных на знаниях. Теоретические аспекты инженерии знаний. Технологии инженерии знаний. Новые тенденции и прикладные аспекты инженерии знаний. Программный инструментарий разработки систем, основанных на знаниях. Пример разработки системы, основанной на знаниях. Интеллектуальные Интернет-технологии.

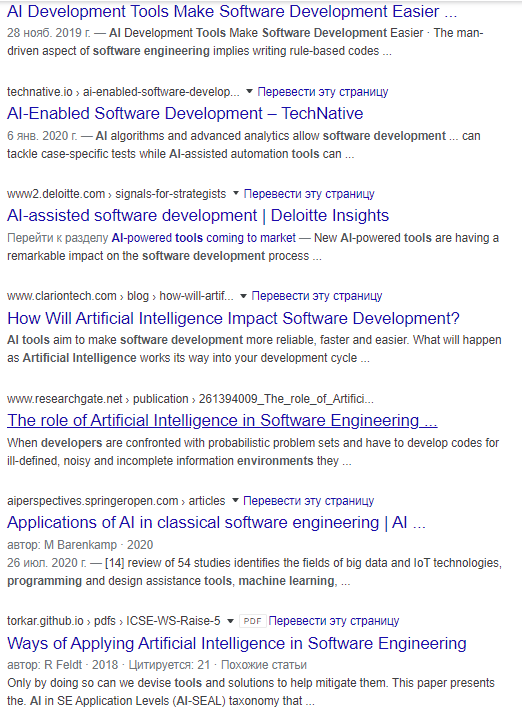
# 2.7. Задача 7

При переводе с русского на английский тема исследования и ключевые слова стали такими: тема – Artificial intelligence software development environments, слова - Artificial intelligence, software development environments, intelligence software development.

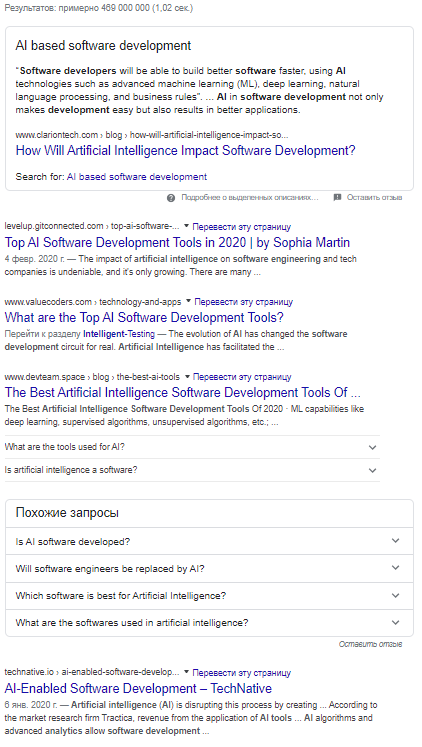
**Итоги поиска в Google:**

Релевантность запроса по теме можно оценить в 8 из 10: сразу выдаются и конкретные примеры, и полезные источники. Пример на скриншоте ниже:





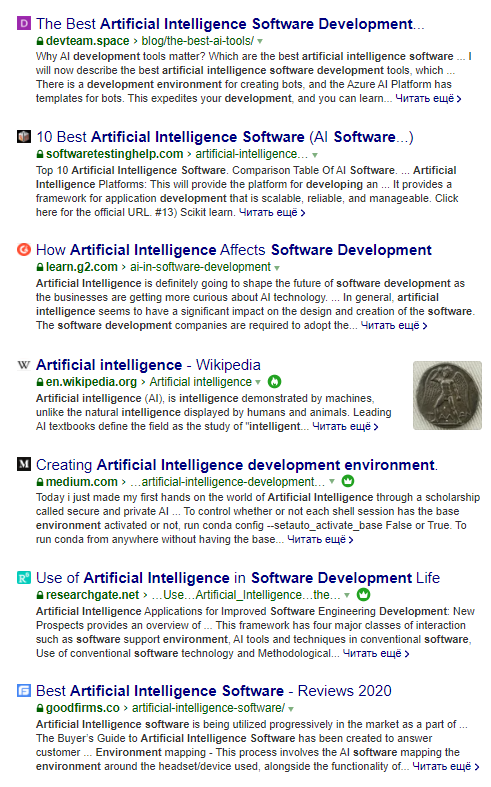
Поиск по ключевым слова выдает похожие ресурсы и в два раза больше источников, которые также могут быть полезными, поэтому оценка релевантности остается 8 из 10:



Расширенный поиск в Google при использовании английских слов выдает релевантные новостные статьи.

**Итоги поиска в Yandex:**

Поиск по теме аналитического отчета сразу оказался достаточно релевантным и предложил новые источники информации:



Поиск по ключевым словам, не сильно отличился от результатов, которые были получены при поиске по теме, и при поиске в другой системе ранее. Были предложены как официальные источники, так и статьи и материалы в соц. сетях. В связи с этим, тяжело давать конкретную оценку поиску в Yandex, по той причине, что она не будет конкретно отличаться от Google.

Расширенный поиск в Yandex за последний месяц не предложил конкретики по теме.

**Общие итоги:**

В выделенных статьях конкретных ключевых слов кроме как «Artificial Intelligence» представлено не было, но при ознакомлении с ними, и другими найденными материалами, очевидно, что также ключевыми для них будут и «Software development environments» и другие слова так или иначе связанные с темой моего аналитического отчета.

На основе использования различных ключевых слов и их изменений, я пришёл к выводу, что существует необходимость изменить ключевые слова, чтобы их суть не была столь схожа. Однако, избавление их от привязки к ИИ не позволит находить нужные статьи и ресурсы по запросам. При этом, запросы на английском выдают куда больше актуальной и полезной информации.

**Работа с наиболее релевантным источником:**

Наиболее релевантным и доступным из найденных материалов я посчитал статью «How Artificial Intelligence Affects Software Development».

# 3. Заключение

По итогам проведенной работы можно сделать ряд выводов:

* Такие поисковые системы как Google и Yandex – не лучшие источники для поиска научных материалов для написания аналитических отчетов или научных статей. Ими, безусловно, можно и стоит пользоваться, однако, используя русский язык, далеко не всегда можно найти полезные материалы, даже с расширенным инструментарием.
* Очень важно владеть английским языком, так как многие темы куда лучше представлены в английском сегменте сети Интернет.
* Подробный разбор текстов и составление различных аннотаций по ним позволяет лучше уложить материал в голове, или же «овладеть им».
* Ключевые слова, будучи использованными в вышеупомянутых поисковых системах, не являются залогом поиска релевантных материалов. Куда проще поиск по словосочетаниям, или используя весь их список разом. Но тут возникает вопрос о корректности длинных ключевых слов, которые несут в себе как можно больше информации по аналитическому отчету.