

# Решение кейса ООО Солюшн

M4 Team

# Состав команды



**Марина Дружинина**

Разработка NLP модели  
извлечение сущностей

Компетенция: разработка программного обеспечения

г. Уфа



@marina\_druzh



**Марат Закиров**

Разработка NLP модели  
разработка интерфейса

Компетенция: разработка программного обеспечения

г. Москва



@mardrake



**Мгер Парунакян**

Парсинг источников

Компетенция: управление разработкой программного обеспечения, производством, дистрибуцией

г. Москва



@MhParun



**Михаил Демин**

Извлечение текста из источников  
менеджмент проекта

Компетенция: экспертиза в фармации и медицине

г. Хабаровск



@MikhailDemin

# Проблематика



На текущий момент отсутствуют легкодоступные обобщенные источники информации о культивировании лекарственных растений. Данный фактор увеличивает затраты на начальном этапе культивирования, делает сложным прогнозирование конечного результата.

**Решение** этой проблемы позволит фермерам и предпринимателям легче принимать управленческие решения и выстраивать бизнес-процессы с учетом особенностей конкретного региона, а также позволит привлечь новых участников к культивированию лекарственных растений на территории Российской Федерации.

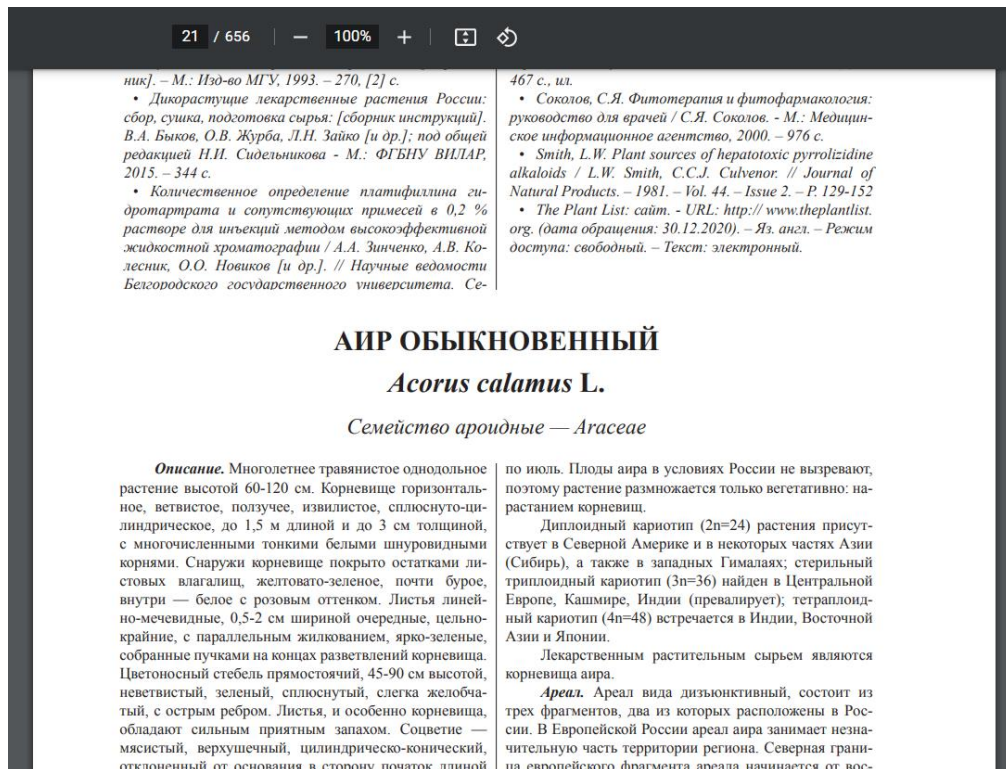


# КРАТКОЕ ОПИСАНИЕ КЕЙСА

- ▶ Необходимо формализовать процесс сбора информации о произрастании лекарственных культур из редких источников, чтобы улучшить поиск необходимых семян для засева определенного типа почв, для увеличения доли отечественного сырья в производстве лекарств.

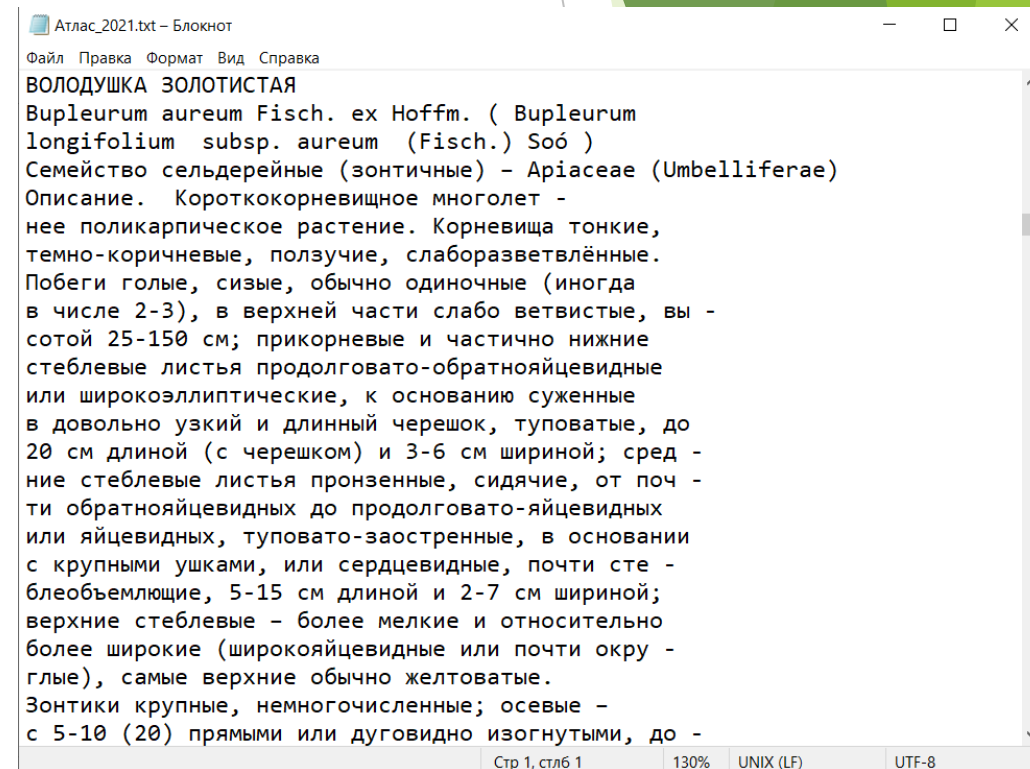


# ИЗВЛЕЧЕНИЕ ТЕКСТА ИЗ PDF



.pdf file

PyPDF2



.txt file

# NLP МОДЕЛЬ

## M4 Team

Атлас\_2021.txt – Блокнот

Файл Правка Формат Вид Справка

- Kataria, H. Phytochemical investigation of medicinal plant *Cnicus wallichii* and *Cnicus benedictus* L. / H. Kataria // Asian Journal of Chemistry. - 1995. - vol.7. - P. 227-228.

- Saghatelyan, A. Flora of Armenia: its composition, analysis and relationships / A. Saghatelyan // Turczaninowia. - 2006. - № 9 (3). - P. 5-47.

- Ulbelen, A. Triterpenic and steroidal compounds of *Cnicus benedictus* L. / A. Ulbelen, T. Berkan / Planta Medica. - 1977. - vol. 31. - P. 375-377.

ВАСИЛЕК ПОСЕВНОЙ (василек синий)

*Cyanus segetum* Hill ( *Centaurea cyanus* L.)

Семейство астровые (сложноцветные) – Asteraceae (Compositae)

Описание. Однолетнее или двулетнее растение высотой 30-80 см, с тонким, стержневым разветвленным корнем и прямым, ветвистым, клочковато-паутиным стеблем. Листья очередные, серо-зеленые, паутинисто-шерстистые; нижние – тройчато- или перистолопастные, реже цельные, черешковые, отмирающие ко времени цветения; выше на стебле листья крупнозубчатые, самые верхние – цельнокрайные, сидячие, линейные. Цветки в соцветиях – корзинках, расположенных на концах стебля и его разветвлений. Обертка корзинок состоит из черепитчато-налегающих друг на друга листочков; наружные и средние из них – эллиптические, по краю беловато-бахромчатые, внутреннее – линейные, желтоватые, на концах перепончатые, цельные или слегка зубчатые. Цветоложе плоское, усеченное длинными щетинками. Краевые цветки

Стр 14,

Den4i kAI  
(BERT)

**БЕДРЕНЕЦ КАМНЕЛОМКА** (бедренец камнеломковый)

*Pimpinella saxifraga* L.

Семейство сельдерейные (зонтичные) – Apiaceae ( Umbelliferae )

Описание . Многолетник, поликарпик, 15-80 см высотой, с вертикальным или косым корневищем, с неутолщенным стержневым корнем. Всё растение обычно сероватое от опушения из мелких волосков. Стебли одиночные, прямые, 2-7 мм в диаметре, ветвистые в верхней части, плотные, тонкобороздчатые, опушенные по всей длине короткими волосками или реже голые, при основании покрыты волокнистыми остатками черешков отмерших листьев. Прикорневые листья многочисленные, с длинными черешками. Черешки с редким опушением, в сечении желобовидные или почти треугольные, плотные или с небольшой полостью в центральной части, с глубокой выемкой или почти плоские с адаксиальной стороны, с периферическими проводящими пучками. Пластинки прикорневых листьев в очертании ланцетные, перистые, опушенные с обеих сторон, 5-15 см длиной, 2,5-7 см шириной, их первичные сегменты крупнозубчатые, округлые или яйцевидные, 15-35 мм длиной, 10-35 мм шириной, тупые, при основании прямые или закругленные; сидячие, реже на коротких черешочках; конечный листочек трёхлопастной или трёхраздельный. Стеблевые листья немногочисленные. быстро

# ИЗВЛЕЧЕНИЕ ИНФОРМАЦИИ ИЗ ДОПОЛНИТЕЛЬНЫХ ИСТОЧНИКОВ

M4 Team

## Красные книги

Списки охраняемых видов растений и лишайников

Ниже перечислены региональные Красные книги, информация из которых публикуется на сайте.

\* жирным шрифтом выделены последние известные версии списков для регионов

**А Б В Г Д Е З И К Л М Н О П Р С Т У Ф Х Ч Э Ю Я**

Регион	Год	Таксоны	Источник	Веб-сайт
<b>А</b>				
1. Республика Адыгея	2011	169	Постановление от 11 октября 2011 г. № 204 о порядке ведения Красной книги Республики Адыгея	<a href="http://oopt.aari.ru">oopt.aari.ru</a> , <a href="http://herba.msu.ru">herba.msu.ru</a>
2. Республика Адыгея	<b>2021</b>	200		<a href="http://docs.cntd.ru">docs.cntd.ru</a>
3. Азербайджанская Республика		138		—

BeautifulSoup



+ Код + Текст

```
[ ] 1 import csv
2
3 with open("region_herb.csv", mode="w", encoding='utf-8') as w_file:
4     file_writer = csv.writer(w_file, delimiter = "\t")
5     file_writer.writerow(["Растение", "Регион"])
6     #сравниваем список растений из файла со списком в словаре
7     for herb in my_herbs_list:
8         for reg in area_herb:
9             if herb in area_herb[reg]:
10                 print(herb + ", " + reg )
11                 file_writer.writerow([herb, reg])
12
```

алтей лекарственный, Брянская ооласть  
алтей лекарственный, Республика Калмыкия  
алтей лекарственный, Курганская область  
алтей лекарственный, Курская область  
алтей лекарственный, Липецкая область  
алтей лекарственный, Омская область  
алтей лекарственный, Пензенская область  
алтей лекарственный, Тамбовская область  
алтей лекарственный, Ульяновская область  
баранец обыкновенный, Республика Башкортостан  
баранец обыкновенный, Брянская область  
баранец обыкновенный, Владимирская область  
баранец обыкновенный, Вологодская область  
баранец обыкновенный, Республика Ингушетия  
баранец обыкновенный, Калужская область  
баранец обыкновенный, Костромская область  
баранец обыкновенный, Курганская область

# РЕЗУЛЬТАТ

M4 Team

result_table.xlsx * x				
	В	С	Д	Е
	Название.	Описание.	Ареал.	Экология.
1	АИР ОБЫКНОВЕННЫЙ	Многолетнее травянистое однодольное растение высотой 60-120 см. Корневище горизонтальное, ветвистое, ползучее, извилистое, сплюснуто-цилиндрическое, до 1,5 м длиной и до 3 см толщиной, с многочисленными тонкими белыми шнуровидными корнями. Снаружи корневище покрыто остатками лиственных влагалищ, желтовато-зеленое, почти бурое, внутри — белое с розовым оттенком. Листья линейно-мечевидные, 0,5-2 см шириной очередные, цельно-крайние, с параллельным жилкованием, ярко-зеленые, собранные пучками на концах разветвлений корневища. Цветоносный стебель прямостоячий, 45-90 см высотой, неветвистый, зеленый, сплюснутый, слегка желобчатый, с острым ребром. Листья, и особенно корневища, обладают сильным приятным запахом. Соцветие — мясистый, верхушечный, цилиндрическо-конический, отклоненный от основания в сторону початок длиной 4-12 см и шириной 0,8-1,5 см. Цветки мелкие, зеленовато-желтые, обоеполые. К початку прилежит длинный, глубокожелобчатый, зеленый, чехлообразный кроющий лист (покрывало). Стебель с кроющим листом почти равен по высоте листьям. Околоцветник невзрачный, простой, из 6 листочков; листочки пленчатые со слегка утолщенными и загнутыми внутрь верхушками. Тычинок 6, расположенных супротивно листочкам околоцветника. Завязь верхняя, трехгнездная, почти шестигранная; рыльце сидячее. Особенности строения указывают на примитивность цветка. Плод — продолговатая, многосемянная, зеленая или красноватая, подсыхающая ягода, снизу с остатками околоцветника. Цветет довольно редко, с конца мая по июль. Плоды аира в условиях России не	Ареал вида дизъюнктивный, состоит из трех фрагментов, два из которых расположены в России. В Европейской России ареал аира занимает незначительную часть территории региона. Северная граница европейского фрагмента ареала начинается от восточного берега Финского залива, идет к югу и, минуя оз. Ильмень, поворачивает на восток, к верховьям рек Мсты и Волги. Севернее известно только одно место нахождения аира — в Южной Карелии. От верховьев Мсты граница ареала аира идет на восток, к среднему течению Клязьмы (во Владимирскую область). Восточная граница ареала резко поворачивает на юг и идет к среднему течению Оки, верховья Дона и выходит на государственную границу с Украиной. Восточнее известны изолированные местонахождения аира в пойме Кудьмы (приток Волги), Рудни (приток Алатыря), в бассейне нижнего течения Суры и на Волге, близ устья Ветлуги. Обособленный участок ареала аира находится в дельте Волги. На юге отмечен изолированный участок 22 ареала в устье Дона. На Кавказе известно изолированное местонахождение аира в плавнях Кубани. На западе граница ареала выходит за пределы государственной границы с Украиной, Белоруссией, Латвией и Эстонией. Азиатский фрагмент ареала аира охватывает ряд районов Западной и Восточной Сибири, а также Дальнего Востока. Западная граница этого фрагмента проходит от города Тары, вверх по течению Иртыша, до государственной границы с Казахстаном и далее отклоняется на восток к Оби, обходит горы Алтая с запада и с севера. Южная граница ареала аира от Оби (южнее Бийска) идет до Енисея (по северным предгорьям Западных Саян) и пересекает эту реку южнее с. Шушенское. Отсюда граница направляется к северу, пересекает западную часть Восточного Саяна, отклоняется	Аир обыкновенный — прибрежно-водное растение, гигрофит, мегатроф. Растет обычно по берегам, погруженным в воду по низинам, стоячим и медленным текущим водам, на илистых и аллювиальных почвах, по пологим и неразмываемым берегам рек, ручьев, стариц, озер, прудов, на заболоченных участках в долинах рек, по мокрым балкам, на автотрофных болотах. В России аир встречается в степной зоне, но известны его отдельные местонахождения и далеко за пределами степей. Поскольку растение снабжено корневищами, за счет них оно разрастается, образуя куртины, иногда довольно большие. Изменение уровня воды в реках и других водоемах, или почвенной влажности, приводит к быстрому исчезновению зарослей аира. В связи с такой чувствительностью к изменению гидрологического режима местообитания аира образуются обычно лентовидные заросли 5-7 м шириной, вытянутые вдоль берегов водоемов. Площадь отдельных массивов с участием аира иногда достигает нескольких тысяч квадратных метров. Однако площадь чистых зарослей аира обычно не превышает нескольких десятков квадратных метров. В смешанных зарослях наряду с аиром встречаются до 15-20 видов растений, но доминируют, как правило, всего 3-5 растений верхнего яруса: тростник обыкновенный ( <i>Phragmites australis</i> (Cav.) Trin. ex Steud.), осока острая ( <i>Carex acuta</i> L.), стрелолист стрелолистный ( <i>Sagittaria sagittifolia</i> L.), жеруха земноводная ( <i>Rorippa amphibia</i> (L.) Bess. ирис желтый ( <i>Iris pseudacorus</i> L.). В среднем и нижнем ярусах произрастают: хвощ болотный ( <i>Equisetum palustre</i> L.), цусак зонтичный ( <i>Butomus umbellatus</i> L.), горец земноводный ( <i>Polygonum amphibum</i> L.), горец перчаточный ( <i>Polygonum hydropiper</i> L.), лапчатка гусиная ( <i>Potentilla anserina</i> L.) и др. В результате
2	АИР ОБЫКНОВЕННЫЙ	Многолетнее травянистое однодольное растение высотой 60-120 см. Корневище горизонтальное, ветвистое, ползучее, извилистое, сплюснуто-цилиндрическое, до 1,5 м длиной и до 3 см толщиной, с многочисленными тонкими белыми шнуровидными корнями. Снаружи корневище покрыто остатками лиственных влагалищ, желтовато-зеленое, почти бурое, внутри — белое с розовым оттенком. Листья линейно-мечевидные, 0,5-2 см шириной очередные, цельно-крайние, с параллельным жилкованием, ярко-зеленые, собранные пучками на концах разветвлений корневища. Цветоносный стебель прямостоячий, 45-90 см высотой, неветвистый, зеленый, сплюснутый, слегка желобчатый, с острым ребром. Листья, и особенно корневища, обладают сильным приятным запахом. Соцветие — мясистый, верхушечный, цилиндрическо-конический, отклоненный от основания в сторону початок длиной 4-12 см и шириной 0,8-1,5 см. Цветки мелкие, зеленовато-желтые, обоеполые. К початку прилежит длинный, глубокожелобчатый, зеленый, чехлообразный кроющий лист (покрывало). Стебель с кроющим листом почти равен по высоте листьям. Околоцветник невзрачный, простой, из 6 листочков; листочки пленчатые со слегка утолщенными и загнутыми внутрь верхушками. Тычинок 6, расположенных супротивно листочкам околоцветника. Завязь верхняя, трехгнездная, почти шестигранная; рыльце сидячее. Особенности строения указывают на примитивность цветка. Плод — продолговатая, многосемянная, зеленая или красноватая, подсыхающая ягода, снизу с остатками околоцветника. Цветет довольно редко, с конца мая по июль. Плоды аира в условиях России не	Ареал вида дизъюнктивный, состоит из трех фрагментов, два из которых расположены в России. В Европейской России ареал аира занимает незначительную часть территории региона. Северная граница европейского фрагмента ареала начинается от восточного берега Финского залива, идет к югу и, минуя оз. Ильмень, поворачивает на восток, к верховьям рек Мсты и Волги. Севернее известно только одно место нахождения аира — в Южной Карелии. От верховьев Мсты граница ареала аира идет на восток, к среднему течению Клязьмы (во Владимирскую область). Восточная граница ареала резко поворачивает на юг и идет к среднему течению Оки, верховья Дона и выходит на государственную границу с Украиной. Восточнее известны изолированные местонахождения аира в пойме Кудьмы (приток Волги), Рудни (приток Алатыря), в бассейне нижнего течения Суры и на Волге, близ устья Ветлуги. Обособленный участок ареала аира находится в дельте Волги. На юге отмечен изолированный участок 22 ареала в устье Дона. На Кавказе известно изолированное местонахождение аира в плавнях Кубани. На западе граница ареала выходит за пределы государственной границы с Украиной, Белоруссией, Латвией и Эстонией. Азиатский фрагмент ареала аира охватывает ряд районов Западной и Восточной Сибири, а также Дальнего Востока. Западная граница этого фрагмента проходит от города Тары, вверх по течению Иртыша, до государственной границы с Казахстаном и далее отклоняется на восток к Оби, обходит горы Алтая с запада и с севера. Южная граница ареала аира от Оби (южнее Бийска) идет до Енисея (по северным предгорьям Западных Саян) и пересекает эту реку южнее с. Шушенское. Отсюда граница направляется к северу, пересекает западную часть Восточного Саяна, отклоняется	Аир обыкновенный — прибрежно-водное растение, гигрофит, мегатроф. Растет обычно по берегам, погруженным в воду по низинам, стоячим и медленным текущим водам, на илистых и аллювиальных почвах, по пологим и неразмываемым берегам рек, ручьев, стариц, озер, прудов, на заболоченных участках в долинах рек, по мокрым балкам, на автотрофных болотах. В России аир встречается в степной зоне, но известны его отдельные местонахождения и далеко за пределами степей. Поскольку растение снабжено корневищами, за счет них оно разрастается, образуя куртины, иногда довольно большие. Изменение уровня воды в реках и других водоемах, или почвенной влажности, приводит к быстрому исчезновению зарослей аира. В связи с такой чувствительностью к изменению гидрологического режима местообитания аира образуются обычно лентовидные заросли 5-7 м шириной, вытянутые вдоль берегов водоемов. Площадь отдельных массивов с участием аира иногда достигает нескольких тысяч квадратных метров. Однако площадь чистых зарослей аира обычно не превышает нескольких десятков квадратных метров. В смешанных зарослях наряду с аиром встречаются до 15-20 видов растений, но доминируют, как правило, всего 3-5 растений верхнего яруса: тростник обыкновенный ( <i>Phragmites australis</i> (Cav.) Trin. ex Steud.), осока острая ( <i>Carex acuta</i> L.), стрелолист стрелолистный ( <i>Sagittaria sagittifolia</i> L.), жеруха земноводная ( <i>Rorippa amphibia</i> (L.) Bess. ирис желтый ( <i>Iris pseudacorus</i> L.). В среднем и нижнем ярусах произрастают: хвощ болотный ( <i>Equisetum palustre</i> L.), цусак зонтичный ( <i>Butomus umbellatus</i> L.), горец земноводный ( <i>Polygonum amphibum</i> L.), горец перчаточный ( <i>Polygonum hydropiper</i> L.), лапчатка гусиная ( <i>Potentilla anserina</i> L.) и др. В результате



# ПОЛЬЗОВАТЕЛЬСКИЙ ИНТЕРФЕЙС M4 Team

To create a public link, set `share=True` in `launch()`.  
Running on <https://localhost:7860/>

region

Краснодарский край

Очистить

Исполнить

output

Unnamed: 0 ▲	Название. ▲	Описание.
22	БЕССМЕРТНИК ПЕСЧАНЫЙ	Травянистый многолетник с войл
70	ГОРИЦВЕТ ВЕСЕННИЙ	Травянистый многолетник высото
185	ПРУТНЯК ОБЫКНОВЕННЫЙ	Невысокое дерево или высокий к
192	САБЕЛЬНИК БОЛОТНЫЙ	Многолетнее травянистое растен
210	СМОКОВНИЦА ОБЫКНОВЕННАЯ	Листопадное многолетнее растен

Пометить



# СПАСИБО ЗА ВНИМАНИЕ!

Ссылка на репозиторий:

