



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования «Московский государственный технический
университет имени Н.Э. Баумана (национальный исследовательский
университет)» (МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления» _____

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии» _____

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ НА ТЕМУ:

Студент: ИУ7-12М
(группа)

_____ Фурдик Н.О.
(подпись, дата) (Фамилия И.О.)

Руководитель

_____ Клышинский Э.С.
(подпись, дата) (Фамилия И.О.)

Москва, 2021 г.

Оглавление

ВВЕДЕНИЕ	3
1. Парсинг сайтов	6
1.1 Библиотека BeautifulSoup4	6
2. Законотворчество	7
3. Описание входных данных	12
3.1 Утилита tesseract	13
ЗАКЛЮЧЕНИЕ	15
СПИСОК ЛИТЕРАТУРЫ	16

ВВЕДЕНИЕ

Государственная Дума выполняет те необходимые функции, без которых не реализуем принцип разделения властей на три части: законодательную, судебную и исполнительную. Эти функции, в свою очередь, делятся на законотворческую, контрольную и представительную. Дума является ключевым законодательным органом - ей принадлежит основная роль в законодательном процессе. Однако, такое право Думы ограничено с двух сторон: отвергнуть принятые ею законопроекты может как Совет Федерации, так и президент РФ.

Поскольку функции Думы чрезвычайно важны для корректного функционирования политической системы РФ, актуальной задачей становится проанализировать работу данного государственного аппарата в разрезе различных метрик. В частности, прохождение законопроекта в зависимости от партии субъектов законопроектов, их департамента или, в конце концов, самой тематики законопроекта. Важно при этом учитывать, как именно происходило решение по проекту: сколько занимал времени переход от одного этапа к другому, имели ли место крупные изменения в законопроекте, либо же закон принимался без каких-либо проблем или крупных поправок.

Стоит также отметить, что проведение исследования в данной области является уместным и может быть показательным в преддверии следующих муниципальных выборов депутатов в Москве в 2022 году, которые определяют состав Государственной Думы на ближайшие пять лет. Учитывая тот факт, что большинство законопроектов вносится в Думу самими депутатами, что в конечном итоге негативно сказывается на законодательном процессе и является нетипичной практикой с точки зрения других стран [1], дополнительный анализ процесса выдвижения законопроектов может быть весьма полезной задачей как для лица, имеющего возможность проголосовать за того или иного кандидата, так и для политических организаций, занимающихся исследованиями в данной области.

В исследовании «Новой газеты», посвященном анализу деятельности Государственной Думы за последние два десятка лет, представляются довольно противоречивые и весьма неутешительные данные: так некоторые законопроекты принимались без установленных регламентом дискуссий – за один день в нескольких чтениях, при этом самые «остросоциальные» проекты могли обсуждаться меньше часа (в частности, закон об ЛГБТ-пропаганде, «дадинская» статья и закон о нежелательных организациях осуждались менее часа - около 45 минут [2]).

Как было сказано ранее, основная часть законов все так же принимается депутатами, но в текущем созыве доля законопроектов, вносимых правительством или президентом резко выросла (по сравнению с шестым) – почти в три раза (депутаты принимают до 92–93% правительственных законопроектов и до 99% инициатив президента [2]), что может означать лишь перетягивание основного функционала Думы, законотворческого, на правительство, что, в свою очередь, может негативно сказываться на деятельности данного государственного аппарата в целом.

Дума, являясь органом государственной власти, действует в сфере политики, однако ее основная деятельность должна заключаться, прежде всего, в законотворчестве, а не в различных политических баталиях. Законодательный процесс становится излишне политизированным: принимается немало законов по малозначительным вопросам [1], а также сохраняется стремление идти путем принятия оперативных решений по законам, требующим длительного и основательного обсуждения (например, «обнуление» сроков текущего президента [2]). Таким образом, стремление бежать за событиями в условиях поляризации политических сил в парламенте в угоду качественной законодательной практики лишь негативно сказывается на всем процессе в целом.

Целью работы является разработка аналитического инструмента для анализа эффективности процедур законотворчества.

Задачи, которые потребуется решить в работе:

- 1) проанализировать варианты эффективной обработки и последующего анализа огромных входных данных;
- 2) разработать архитектуру приложения, в том числе алгоритм, позволяющий наилучшим способом выполнить анализ данных;
- 3) спроектировать, реализовать и протестировать ПО;
- 4) исследовать разработанное ПО на эффективность.

Областью применения могут являться политические организации, заинтересованные в исследованиях в данной области, СМИ, а также высшие учебные заведения, имеющие у себя политические направления.

1. Парсинг сайтов

Неотъемлемой частью анализа больших данных является предобработка данных в структурированную форму. Поскольку для текущей работы все данные можно найти на веб-страницах, нужно нормализовать способ их получения. Одним из наиболее удачных вариантов являются инструменты web scraping (парсинга), которые разрабатывались для извлечения и сбора любой открытой информации с веб-сайтов. Эти ресурсы нужны тогда, когда необходимо быстро получить и сохранить в структурированном виде любые данные из интернета.

Такого рода программное обеспечение собирает 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. Инструменты парсинга могут применяться с разными целями и в различных сценариях, поэтому они замечательно подойдут под нашу задачу скачивания законов с официального сайта Государственной Думы.

1.1 Библиотека BeautifulSoup4

BeautifulSoup4 — это библиотека для языка Python, которая применяется для синтаксического анализа документов HTML и XML формата (в том числе может распознавать искаженную разметку - незакрытые теги). В процессе обработки создается парсинг-дерево для всех проанализированных страниц, которое впоследствии можно использовать как инструмент извлечения данных из HTML, что позволяет эффективно парсить веб-страницы.

BeautifulSoup4 была разработана Леонардом Ричардсоном, а также с некоторого момента дополнительно поддерживается Tidelif, платной подпиской на обслуживание с открытым исходным кодом.

Данная библиотека доступна для Python 2.7 и Python 3.

BeautifulSoup4 имеет следующие преимущества:

- низкий порог вхождения, понятный уровень документации;
- устойчивость к HTML ошибкам (возможность распознавать искаженную разметку);
- возможность анализа сложных приложений.

Из недостатков можно выделить:

- невозможность чтения динамического содержимого;
- отсутствие ярко выраженной экосистемы.

2. Законотворчество

Любой принятый законопроект начинается с законодательной инициативы. Список тех, кто может положить начало законопроекту, а также порядок внесения законопроектов в Государственную Думу определены в ст. 104 Конституции РФ. Среди лиц, обладающим данным правом:

- Президент РФ;
- Совет Федерации и его членам;
- депутаты Государственной Думы;
- Правительство РФ;
- законодательные (представительные) органы субъектов РФ;
- Конституционный суд РФ;
- Верховный суд РФ.

Подготовленный к внесению в Государственную Думу законопроект и материалы к нему направляются на имя Председателя ГД. Поступивший законопроект регистрируется в Управлении документационного обеспечения Аппарата ГД в Системе автоматизированного делопроизводства и документооборота ГД. Ему присваивается регистрационный номер, который указывается вместе с наименованием законопроекта в течение всего периода прохождения законопроекта в Государственной Думе.

Одновременно в Системе обеспечения законодательной деятельности (СОЗД) на законопроект заводится электронная регистрационная карта, в которой фиксируются дата и время поступления документа, список тех, кто его внес, сведения о прохождении законопроекта в ГД, об одобрении Государственной Думой федерального конституционного закона или о принятии федерального закона и о рассмотрении соответствующего закона Советом Федерации и Президентом. То есть в СОЗД можно отследить прохождение законопроектом всех этапов.

Поступившие документы Председатель ГД направляет в профильный комитет, который проверяет их на соответствие установленным требованиям.

Затем прошедший проверку документ направляется в Совет Государственной Думы. В состав Совета входят Председатель ГД, Первые заместители Председателя ГД, заместители Председателя ГД, руководители фракций.

Следующие ответственные комитеты могут направить проект закона для предоставления отзывов, замечаний и предложений:

- Совет Государственной Думы.
- Президент РФ;
- Счетная палата;
- Правительство РФ;
- Общественная палата;
- Конституционный суд;
- Верховный суд РФ.

Совет может вернуть законопроект профильному комитету или инициатору, но если все в порядке, Совет назначает ответственными за документ один или несколько комитетов ГД и включает его в примерную программу работы Государственной Думы.

Далее основная работа над законопроектом идет в ответственном комитете. При этом для работы над документом могут привлекаться эксперты,

представители профессиональных и научных сообществ, создаваться различные рабочие группы, в которых ведется анализ предложений по совершенствованию законопроекта, проводятся парламентские слушания, круглые столы и т. д.

После того, как ответственный комитет проведет заседание и рекомендует депутатам поддержать или отклонить законопроект, начинается рассмотрение документа на пленарных заседаниях.

В целом, весь процесс прохождения законопроекта в Государственной Думе, Совете Федерации и у Президента можно представить на рисунке 1.

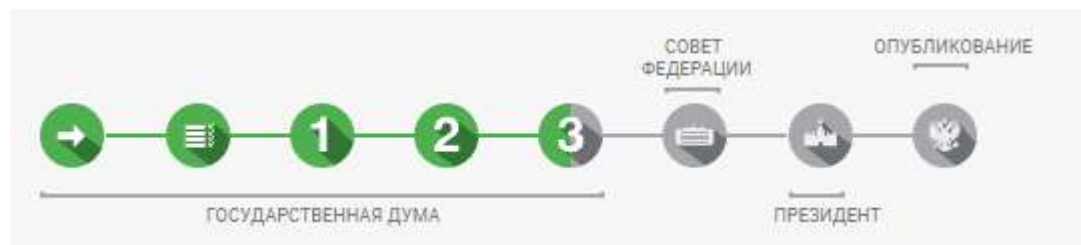


Рисунок 1. Стадии рассмотрения законопроекта (Пример законопроекта, в данный момент рассматриваемого на третьем чтении)

Как видно из рисунка, в Государственной Думе законопроект проходит три чтения.

Первое чтение

В первом чтении рассматривается концепция законопроекта, его актуальность и практическая значимость, дается оценка соответствия основных положений законопроекта Конституции. Обычно во время пленарного заседания с докладом выступает субъект права законодательной инициативы, внесшего законопроект, или его представитель. С содокладом выступает представитель ответственного комитета.

Поправки к законопроекту, принятому или одобренному в первом чтении, вносятся в ответственный комитет в виде:

- изменения редакции статей;

- дополнения законопроекта конкретными статьями;
- предложений об исключении конкретных слов, пунктов, частей или статей законопроекта.

Поправки не должны противоречить концепции законопроекта.

Решение о принятии или об одобрении законопроекта в первом чтении оформляется соответствующим постановлением ГД. Если по итогам голосования предложение о принятии проекта в первом чтении не набрало необходимого числа голосов, то он считается отклоненным без дополнительного голосования. Решение об отклонении законопроекта также оформляется соответствующим постановлением ГД. Отклоненный документ дальнейшему рассмотрению не подлежит.

Второе чтение

После того, как ответственный комитет на заседании вновь обсудит документ и поступившие к нему поправки, приходит время второго чтения. На этом этапе законопроект обсуждается детально, постатейно, с различными вариантами поправок. Их ответственный комитет представляет в виде таблиц:

- таблица поправок, рекомендуемых к принятию и включенных в текст законопроекта;
- таблица поправок, рекомендуемых к отклонению;
- таблица поправок, по которым не было принято решений.

Принятый во втором чтении документ направляется в ответственный комитет для устранения возможных внутренних противоречий, установления правильной взаимосвязи статей и для редакционной правки, необходимой в связи с изменениями, внесенными в текст законопроекта. После этого проект вместе с заключением Правового управления Аппарата ГД представляется ответственным комитетом в Совет ГД. Тот включает его в проект порядка работы Государственной Думы для рассмотрения в третьем чтении.

В СМИ второе чтение часто называют основным — это связано с тем, что именно в этом чтении законопроект приобретает свой окончательный вид и обсуждается на пленарном заседании наиболее подробно.

Третье чтение

В третьем чтении законопроект принимается в целом — допустимы лишь редакционные уточнения. Рассмотрение законопроекта в третьем чтении начинается с доклада представителя ответственного комитета. Перед голосованием допускаются выступления представителей фракций.

В исключительных случаях по требованию фракций председательствующий обязан поставить на голосование вопрос о возвращении законопроекта к процедуре второго чтения.

Одобрённые ГД федеральные конституционные законы и принятые федеральные законы с соответствующими постановлениями Государственной Думы, стенограммами пленарных заседаний, заключениями Правительства и другими необходимыми материалами оформляются ответственным комитетом совместно с Аппаратом ГД и в течение пяти дней передаются Государственной Думой на рассмотрение Совета Федерации. Обязательному рассмотрению в Совете Федерации подлежат:

- федеральные конституционные законы;
- законы о поправках в Конституцию;
- финансовые законы;
- законы о ратификации и денонсации международных договоров РФ;
- законы по вопросам войны и мира, статуса и защиты государственной границы РФ.

Остальные законы, поступившие из ГД, Совет Федерации может не рассматривать и не обсуждать — это будет означать согласие с принятием закона.

Последний этапom является подписание закона у Президента. Принятый федеральный закон в течение пяти дней направляется Президенту РФ для подписания и обнародования, и в течение 14 дней глава государства подписывает федеральный закон и обнародует его.

Если Президент в течение двух недель с момента поступления федерального закона отклонит его, то Государственная Дума и Совет Федерации в установленном Конституцией порядке вновь рассматривают закон. Если при повторном рассмотрении федеральный закон будет одобрен в ранее принятой редакции большинством, не менее двух третей голосов от общего числа членов Совета Федерации и депутатов Государственной Думы, он подлежит подписанию Президентом в течение семи дней и обнародованию.

Официальным опубликованием считается первая публикация полного текста закона:

- в «Парламентской газете»;
- в «Российской газете»;
- в «Собрании законодательства Российской Федерации»;
- на Официальном интернет-портале правовой информации.

Федеральные конституционные законы и федеральные законы подлежат официальному опубликованию в течение семи дней после момента их подписания Президентом.

Закон вступает в силу через 10 дней после официального опубликования, если самими законами не установлено другое.

3. Описание входных данных

С помощью BeautifulSoup4 происходит парсинг страниц сайта. Со страницы с законом выкачиваются файлы с текстом закона, пояснительная записка, а также различного рода метаданные, в том числе:

- субъект права законодательной инициативы;

- форма законопроекта;
- ответственный комитет;
- отрасль законодательства;
- тематический блок законопроектов;
- созыв, в котором был принят законопроект;
- год принятия.

Также со страницы с ЭРК по хронологии выкачиваются даты, события, строится последовательность событий для корректного последующего анализа.

Важно учитывать расширение файлов, которые скачиваются, так как каждый обрабатывается по-своему.

Так, .pdf обрабатывается с помощью утилиты tesseract, .doc – с помощью библиотеки Win32.App и т.д. В конечном итоге получается .txt файл, который является оптимальным решением с точки зрения занимаемого дискового пространства и обработки.

3.1 Утилита tesseract

Tesseract является одним из самых мощных механизмов оптического распознавания символов для различных операционных систем; бесплатное ПО, выпущенное по лицензии Apache.

В 2006 году Tesseract считался одним из наиболее точных доступных механизмов распознавания с открытым исходным кодом.

Tesseract имеет огромную область применения, в частности:

- распознавание паспортов в аэропортах;
- автоматизация ввода данных;
- распознавание номерных знаков;
- извлечение информации о визитной карточке;
- преобразование рукописных документов в электронные изображения;

- создание PDF-файлов с возможностью поиска;
- создание звуковых файлов (преобразование текста в аудио).

Tesseract является утилитой, обученной на примерах, что делает его эффективным и гибким под разные задачи.

Примерный ход работы программы представлен на рисунке 2.

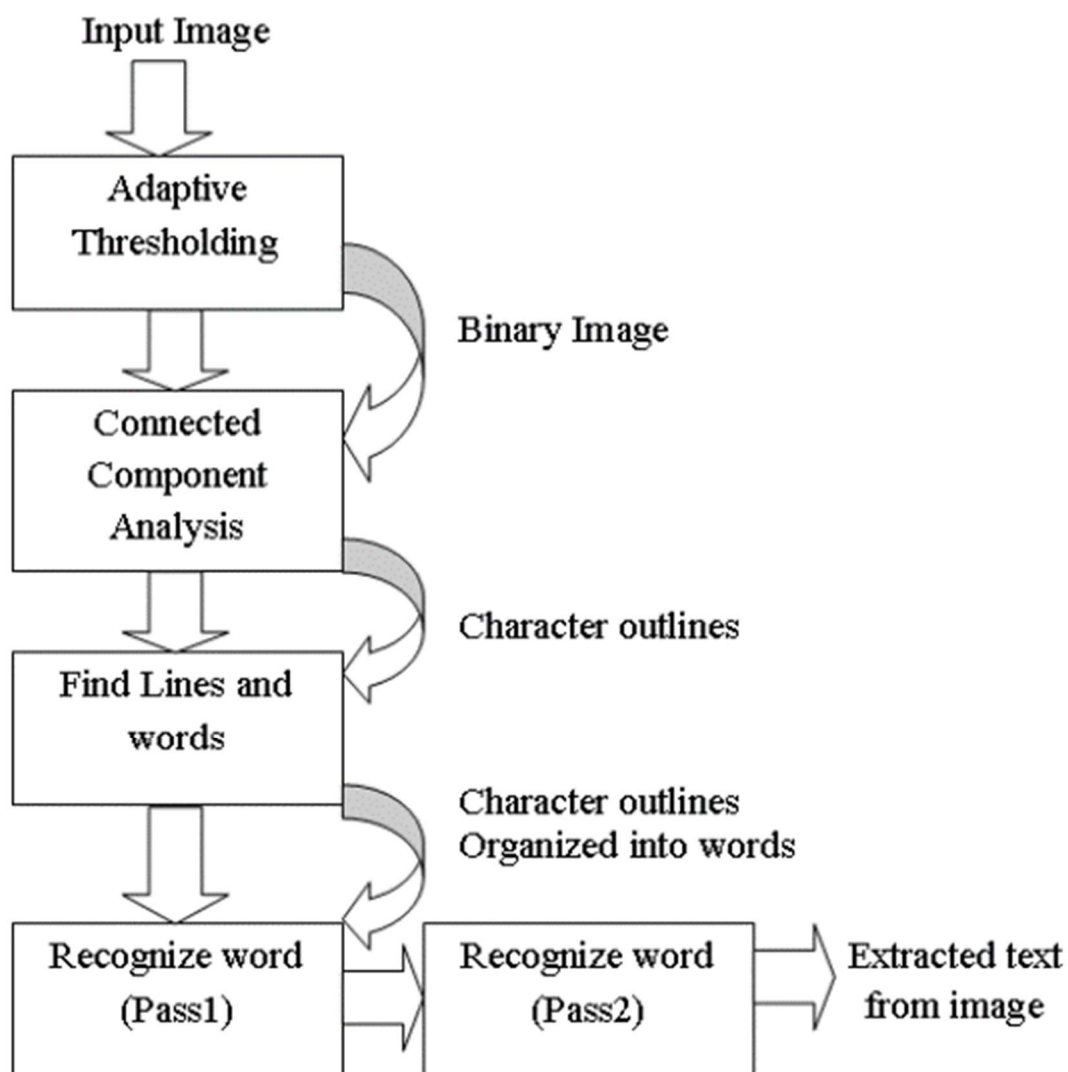


Рисунок 2. Принцип работы tesseract

ЗАКЛЮЧЕНИЕ

Таким образом, в результате данного исследования были достигнуты поставленные цели:

- была проанализирована предметная область;
- были описаны входные данные.

СПИСОК ЛИТЕРАТУРЫ

1. Государственная Дума в политической системе РФ. [Электронный ресурс]. URL: <http://www.hist.msu.ru/Science/Conf/lomweb01/gudev.htm> (дата обращения 10.02.2022)
2. Что вы принимаете? [Электронный ресурс]. URL: <https://novayagazeta.ru/articles/2021/09/15/chto-vy-prinimaete> (дата обращения 10.02.2022)
3. BeautifulSoup Documentation [Электронный ресурс]. URL: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/> (дата обращения 10.02.2022)
4. Как инициатива становится законом [Электронный ресурс]. URL: <http://duma.gov.ru/news/28305/> (дата обращения 10.02.2022)
5. IMPROVING THE EFFICIENCY OF TESSERACT OCR ENGINE [Электронный ресурс]. URL: https://scholarworks.sjsu.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1416&context=etd_projects (дата обращения 10.02.2022)