

Asesor para "navegar" normativas (LLM + DB normas)

Recuperación y Minería de Texto 2024

Bettachini, Víctor; Bringas, Mauro; Nesterov, Igor; Costa Vila, Hernán

Contexto

Industria nuclear en Argentina:

1. Producción de energía limpia y confiable: La energía nuclear es una fuente de energía baja en emisiones de gases de efecto invernadero. Las plantas nucleares, como Atucha I, II y Embalse, son parte esencial del mix energético argentino.
2. Aplicaciones médicas y científicas: La tecnología nuclear médica (radioisótopos para tratamientos oncológicos) e industrial, contribuyen al desarrollo económico.
3. Exportación de tecnología: Argentina exporta reactores de investigación y servicios nucleares a otros países, lo que diversifica su economía y refuerza su posicionamiento en el escenario global.

Contexto

- La Autoridad Regulatoria Nuclear (ARN) se crea en 1994 para oficial de ente regulador independiente de organismos, como la Comisión Nacional de Energía Atómica (CNEA), que son los que trabajan en el área.
- Dada su relativa juventud, carece de un conjunto abarcativo de normativas fruto de experiencia en tarea regulatoria, como organismos de función idéntica en otros países con más historia, e.g. la Nuclear Regulatory Commission de los E.U.A. data de 1975.
- Posibles aspectos a desarrollar son
 - Herramientas para "orientar" a usuarios no científicos de tecnología nuclear.
 - Normativa de actividades nucleares que aún no se desarrollan en el país.

¿Qué datos vamos a usar?

- Utilizaremos normas y guías específicas desarrolladas por la Autoridad Regulatoria Nuclear (ARN).
- Estas se encuentran disponibles en el sitio argentina.gob.ar, y pueden descargarse en formato PDF.



Normas Regulatorias



Guías Regulatorias



Scrapping del texto en PDFs

- 64 Normas: "establecen los requisitos para las instalaciones y prácticas, y para el personal que se desempeña en ellas"
- 10 Guías: "contienen recomendaciones para cumplir con los requisitos de las normas regulatorias vinculadas"

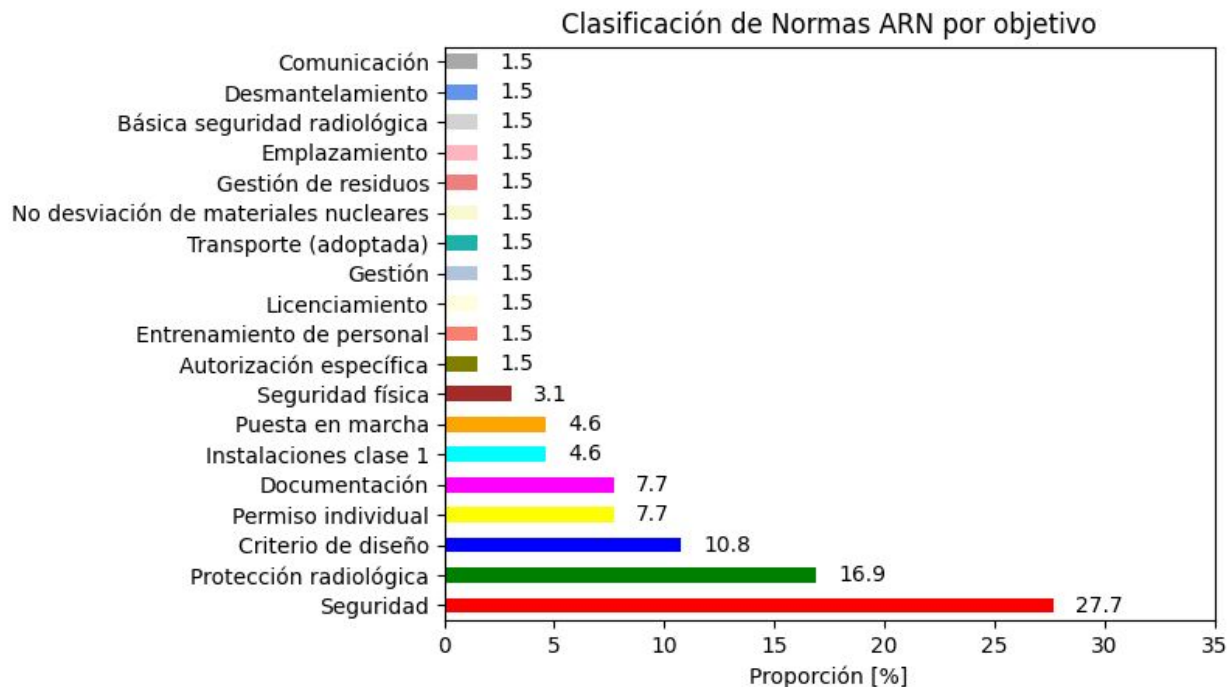
Visto el tamaño del corpus, se utilizó una herramienta online (<https://pdftotext.com/es/>) para obtener texto plano a partir de estos documentos, que luego procesamos en Python.

Clasificación de normas ARN

Pedimos a un profesional de ARN que les clasificara por objetivo

Normas		Objetivo
0	0-0-1_r2	Instalaciones clase 1
1	0-11-1_r3	Instalaciones clase 1
2	0-11-2_r2	Autorización específica
3	0-11-3_r1	Entrenamiento de personal
4	0-11-4_r0	Licenciamiento
...
59	8-11-3_r0	Permiso individual
60	8-2-1_r0	Seguridad
61	8-2-2_r1	Seguridad
62	8-2-3_r3	Seguridad
63	8-2-4_r1	Seguridad

64 rows x 2 columns

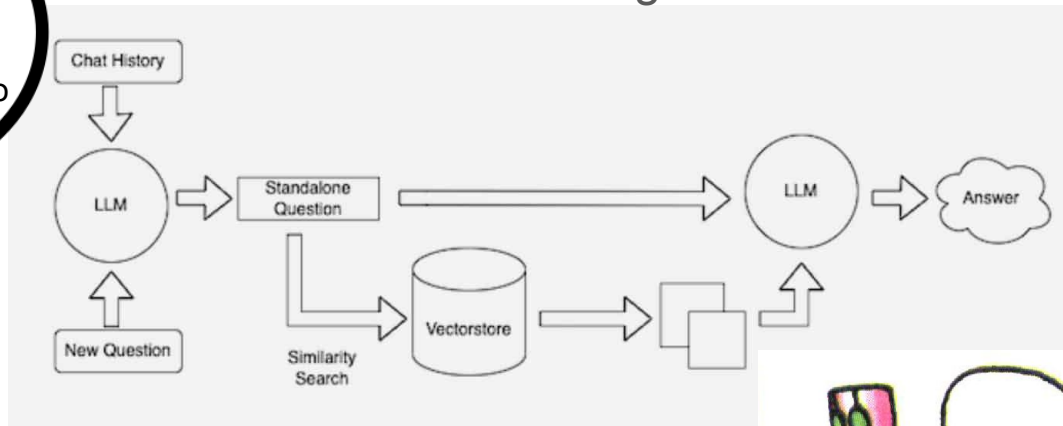


Objetivo 1: RAG para consultar normativa nuclear

Quiero fabricar radiofármacos para diagnóstico con un acelerador de protones. ¿Qué debo tener en cuenta?



Retrieval Augmented Generation



Lo primero es calcular las máximas actividades de los radionucleidos involucrados en cada paso del proceso.

¿Por qué está bueno?

- Esto puede servir como apoyo de emprendedores que quieran introducirse al rubro nuclear.
- Les permitirá tener en cuenta los factores determinantes para cumplir con la normativa vigente.



Grafo dirigido de normas

- Primer enfoque: buscar referencias directas entre normas

NORMA AR 0.11.2. REQUERIMIENTOS DE APTITUD PSICOFÍSICA PARA AUTORIZACIONES ESPECÍFICAS

- Estabilidad emocional y psíquica, así como capacidad de reacción;
- Agudeza de los sentidos que permita una capacidad normal de recepción de la información hablada y escrita, y de toda otra información audible, visible o palpable en la forma en que la proporcionen los aparatos o dispositivos previstos para ello, así como una capacidad de expresión que permita una comunicación rápida y segura;
- Capacidad física, de potencia motora, de campo de movimientos y de destreza de movimientos, tal que le permitan ejecutar adecuadamente las tareas asignadas.

24. Para otorgar el certificado de aptitud psicofísica el médico examinador debe, en el grado requerido por el profesiograma psicofísico de la función especificada, verificar el cumplimiento de las condiciones que correspondan. En la Guía Regulatoria GR 3/AR 0.11.2. se describen condiciones que pueden ser seguidas con este propósito.

13. El sistema de protección debe ser diseñado de manera tal que se asegure que las fallas de la instalación no causen la pérdida de funciones de protección y que se cumpla la Norma AR 3.1.3. "Criterios Radiológicos Relativos a Accidentes en Reactores Nucleares de Potencia".

Objetivos alternativos

Objetivo 2: determinar si algo está normado en otro país pero no por ARN

- Tomar normativa de otro país, e.g. Nuclear Regulatory Commission
- Traducir normas del idioma original al español.
- Correlacionar normas extranjeras y argentinas.



Objetivo 3: asesor para la escritura de nuevas normas

- Sugiera una redacción con la misma terminología para evitar re-definiciones
- Siga el glosario de la Agencia Internacional de Energía Nuclear
- Advierta reiteraciones o contradicciones con normativas ARN e IAEA



El dataset de ARN es "pequeño"

- La cantidad de normas / guías es abarcable para un lector. ¿Vale la pena diseñar un asesor basado en LLM?
- Utilizaremos este conjunto acotado para hacer una prueba de concepto y verificar que las respuestas “son de buena calidad”. ¿Diseñar una métrica?
- Nos proponemos reutilizar la estrategia para un conjunto de normas de mayor tamaño...

Fármacos, alimentos y tec. médica

Legislación

Normativa vigente que rige nuestra actividad regulatoria.

Disposiciones

Consultá los actos dictados por esta Administración.

Consultá toda la legislación vigente de la ANMAT:

Normativa general

Prohibición de uso y/o comercialización

Clausuras e inhabilitaciones de establecimientos

Buenas Prácticas Regulatorias y Buenas Prácticas de Reliance

Circulares



Argentina.gob.ar

Buscá en el sitio



Mi Argentina

Ministerio de Salud /

anmat

En la Administración Nacional de Medicamentos, Alimentos y Tecnología Médica protegemos a la población garantizando que los productos para la salud

Web Scrapping ANMAT

"Recorrer" múltiples botones → biblioteca Selenium

Medicamentos	Registro de Especialidades Medicinales	Disposición N° 8227/2024	2024	Ampliarse los alcances de la Disposición ANMAT N° 4171/24 para todas las especialidades medicinales inscriptas en el Registro de Especialidades Medicinales (REM) de esta Administración Nacional que contengan los siguientes Ingredientes Farmacéuticos Activos (IFAs), a saber: OMEPRAZOL como monodroga, hasta la concentración de 40 mg, ESOMEPRAZOL como monodroga, hasta la concentración de 40 mg; LANSOPRAZOL, como monodroga, hasta la concentración de 30 mg y PANTOPRAZOL, como monodroga, hasta la concentración de 40 mg, a partir de la entrada en vigencia de la presente disposición	12/09/2024	descargar
Normativa general	Normativa general	Decreto N° 94/23	2023	Designase en el cargo de Administradora Nacional de la ADMINISTRACIÓN NACIONAL DE MEDICAMENTOS, ALIMENTOS Y TECNOLOGÍA MÉDICA (ANMAT), organismo descentralizado actuante en la órbita del MINISTERIO DE SALUD, a la doctora Nélida Agustina BISIO (D.N.I. N° 11.770.609).	27/01/2024	descargar
Normativa general	Normativa general	Disposición N° 10867/23	2023	Sustitúyese el texto del ANEXO I (IF-2019-06341144-APN-DER#ANMAT) de la Disposición ANMAT N° 4729/2019, por el texto que registrado como IF-2023-124930809-APN-DLEIAER#ANMAT - "GUÍA DE RECOMENDACIONES PARA UN MENÚ LIBRE DE GLUTEN SEGURO" y su correspondiente lista de verificación, forma parte integrante de la presente disposición.	13/12/2023	descargar

```
# es una pagina dinámica, entonces utilizamos Selenium
driver = webdriver.Chrome()
driver.get("https://www.argentina.gob.ar/anmat/normativa-general")

pdf_links = []

for i in range(36):
    time.sleep(5)

#guardar botones
buttons = driver.find_elements(By.CLASS_NAME, 'btn.btn-primary.btn-sm.margin-btn')
pdf_links.extend([button.get_attribute('href') for button in buttons])

#la pagina siguiente
next_button = driver.find_element(By.XPATH, '//a[@aria-label="Ir a la página siguiente"]')

#no hay mas paginas
if 'disabled' in next_button.get_attribute('class'):
    # break

next_button.click()

driver.quit()
```

Webscrapping ANMAT

No se trata solo de PDFs, también hay textos planos publicados:

- PDFs, raíces de nombres de archivos
 - circular_
 - anmat_circular_
 - anmat_disposicion_
- HTML, *ídem*.
 - circular-
 - decreto-
 - disposici%C3%B3n-
- HTML con archivo PDF adjunto.
 - algunos de los enlaces llevan al sitio boletinoficial.gob.ar, donde ciertos textos también incluyen archivos PDF con las normas relacionadas.
- Otros...

