



Caso de uso Repositorio CIMMYT

Gerardo Flores Petlalcaco
Soporte Técnico La Referencia



Un poco de contexto

- Los crawlers nacieron cuando internet comenzó a ser más grande.
- Mantuvieron bajo perfil hasta la creación de la IA.
- Buscan servicios que tengan información de calidad.
- Los repositorios son buena fuente de información.
 - Conocimiento libre.
 - Bien estructurado



Nuestra experiencia

- Venimos experimentando desde dos años a la fecha tráfico nocivo hacia el repositorio.
- Cambios internos provocaron la migración hacia servicios alternativos.
 - Las herramientas no cumplían con los requerimientos del repositorio en cuanto a manejo de tráfico intenso.



Problemas

- Intermitencias en el servicio.
- Imposibilidad de que usuarios legítimos consulten datos.

A nivel institucional

- Detienen la generación de conocimiento.
- Daño en la reputación de la institución.

Medidas paliativas

- Aumentar los recursos del servidor.
- Crear una estrategia de bloqueo por agentes.
- Bloqueo por IP.
- Restringir países enteros.



Todos estas medidas son temporales porque pueden ser evadidas y volver a tirar los sitios



Anubis



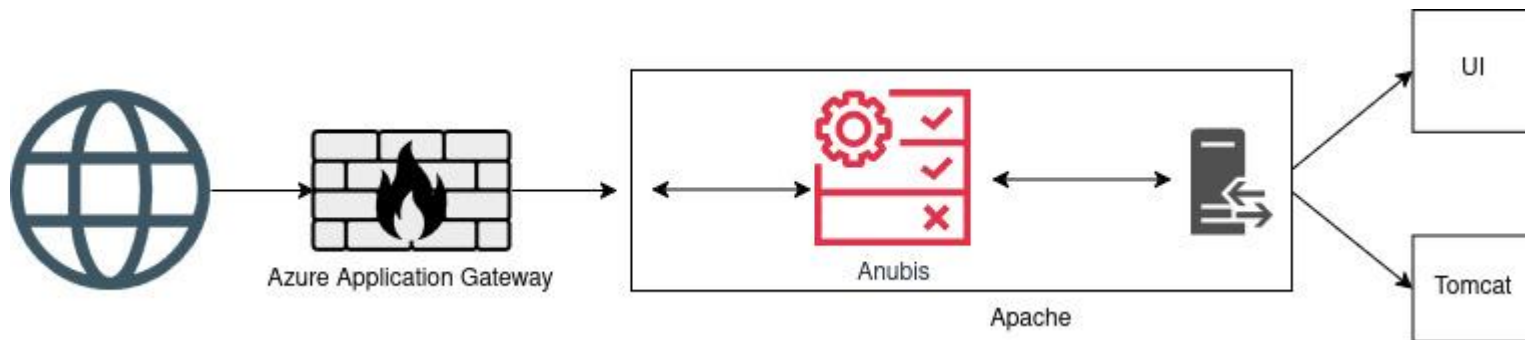
- Un firewall de uso libre.
- Adaptable mediante políticas personalizadas
- Usa detección por ratelimit y heurísticas de comportamiento



Retos

- Integrarse con la solución actual de WAF institucional.
- Integración la autenticación mediante protocolo SAML (Shibboleth).
- Mantener la integridad de las estadísticas.

Arquitectura Actual





Trabajo a futuro

- Actualizar el repositorio a una versión con mejor manejo de recursos.
- Cambiar de servidor Web - Apache -> NginX

Documentos de apoyo y contacto



SOPORTE TÉCNICO

Escanea para abrir una solicitud

<https://soporte-dspace.lareferencia.info/es/home>



AGENTE IA NOTEBOOK

Escanea para probar el Agente IA

<https://notebook.google.com/notebooks/87b5ea36-bdbc-4ba2-b156-d38d9c1a7cae/preview>



CÓDIGO QR DISCORD

Escanea para unirte al canal oficial

<https://discord.com/invite/GQzvHREzNy>