1. Introducción

# 1.1 Propósito

La base de datos de conversación contiene una lista de palabras, ordenadas según su frecuencia como tema de conversación, que se usarán durante la interacción con el usuario.

Además clasifica las palabras según su gramática y semántica para poder estructurar las oraciones de una manera coherente.

# 1.2 Alcance y enfoque

El proyecto debe enfocarse en un flujo de los datos de manera rápida y una administración eficiente para el tiempo de respuesta del chatbot.

El alcance de los datos se limita a los tópicos aprobados por el cliente. La versión inicial del aplicativo será en idioma español debido al algoritmo implementado para la formación de oraciones.

Por tanto quedan fuera del alcance aquellos datos que no se incluyan específicamente además de palabras que no sean del idioma principal.

# 1.3 Descripción general del sistema

La aplicación que utilizará la base de datos es un chatbot enfocado a la interacción con niños en el espectro autista. Por ello el tiempo de respuesta debe ser rápido y las respuestas coherentes según el tópico que se esté tratando.

| **Resumen del sistema** | **Detalles** |
| --- | --- |
| Patrocinador del Proyecto | Ministerio de Educación |
| Nombre del sistema | Picto |
| Tipo de sistema | Aplicación principal |
| Estado operacional | En proceso de modificación |
| Condiciones especiales | El proyecto debe ser desarrollado en seis meses.  Todas y cada una de las herramientas deben ser pensadas para ser llamativas e interesantes para el usuario final. |

# 1.4 Siglas y abreviaturas.

| **Acrónimo/Abreviación** | **Significado** |
| --- | --- |
| CCNT | Chatbot para la comunicación con niños con TEA |
| BDL | Base de datos del lenguaje |
| ABD | Administrador de la base de datos |
| POC | Punto de contacto organizacional |

# 1.5 Puntos de contacto

| **Rol** | **Nombre** | **Correo electrónico** | **Teléfono** |
| --- | --- | --- | --- |
| Lider de proyecto | Keler | keler.modesto@unmsm.edu.oe | 989893902 |
| Analista de datos | Héctor Rodrigo Huaripata Huaccha | hector.huaripata@unmsm.edu.pe | 910178406 |
| Analista de datos | Daniel Triveno Ruffner | daniel.triveno@unmsm.edu.pe | 984300461 |
|  |  |  |  |

## 1.5.1 Propietarios de datos

| **Tipo de datos** | **Nombre del PCO** | **Correo electrónico** | **Teléfono** |
| --- | --- | --- | --- |
| General | Centro de Investigación en Conducta y Cognición Comparada | fcabrera@cencar.udg.mx | +52 (392) 925 9400 |
| Deporte | Ministerio de Educación | educacionlima@drelm.gob.pe | +51 (01) 615-5800 |
| Escuela | Ministerio de Educación | eanahua@drelm.gob.pe | +51 (01) 615-5800 |

2 Resumen del sistema

La base de datos almacenará las palabras más frecuentes en el idioma, además de otras elegidas según distintos tópicos. Su funcionamiento se enfoca en el tiempo de respuesta a las querys ya que el chatbot que las necesita debe dar una respuesta inmediata.

2.1 Información del sistema

2.1.1 Configuración del sistema de gestión de bases de datos

| **Hardware** | **Versión** | **Comentarios** |
| --- | --- | --- |
| Intel Core i7-10750H |  |  |
| NVIDIA GeForce GTX 1660 Ti |  |  |
| DRAM DDR4 16GB |  |  |

2.1.2 Utilidades de software de base de datos

El software principal para el funcionamiento de las bases de datos.

| **Producto** | **Versión** | **Comentarios** |
| --- | --- | --- |
| MySQL | 8.0 | Gestión de la base de datos |
| IIS DB Manager | 1.0 | Software para el soporte en windows de la base de datos. |

2.1.3 Software de Soporte

Software utilizado para dar mantenimiento u optimizar las funciones de las bases de datos.

| **Producto** | **Versión** | **Propósito** |
| --- | --- | --- |
| SQL Tuning | En línea | Realizar pruebas durante el desarrollo. |
| Oracle Database | 12.1.0.1 | Software para el almacenamiento de las bases de datos. |

| **Documento** | **Versión** | **Sección/Página** |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

2.1.4 Seguridad

Para asegurar los datos en el sistema, se diseñarán las bases de datos teniendo en cuenta las pautas redactadas en ISO 27001. Además, se seguirán las recomendaciones de seguridad de la OWASP.

2.2 Arquitectura

Debido a la extensión del proyecto, es conveniente utilizar los formatos de archivo estándar.

2.2.1 Arquitectura de software

Los datos tendrán los siguientes atributos:

* Palabra clave : String
* Clasificación gramatical : enum
* Frecuencia de uso : float
* clasificación semántica : float[ ]

2.2.2 Almacenes de datos

Los datos registrados se guardarán en 3 tablas:

* Tabla de palabras
* Tabla de frases
* Tabla de secuencias

**3 Decisiones de diseño de bases de datos**

3.1 Suposiciones

| **Ref #** | **Suposición** | **Impacto** |
| --- | --- | --- |
| #1 | La información en la base de datos será sellada después de cada actualización. | El usuario final no podrá modificar la base de datos. |
| #2 | El software es privado. Todos los datos deben ser asegurados y su procesamiento debe ser cerrado. | El diseño de la base debe ser enfocado en la seguridad y privacidad de los datos. |

3.2 Problemas

En situación de conflicto se deben tomar en cuenta los lineamientos del proyecto mayor, sin embargo, de aún así no resolverse, se delega la decisión al director de proyecto.

**4 funciones administrativas de la base de datos**

4.1 Responsabilidad

| **Rol** | **Nombre** | **Responsabilidad** | **Correo electrónico** |
| --- | --- | --- | --- |
| Administrador de la base de datos |  | Manejar, administrar y mantener la base de datos |  |
| Administrador del sistema |  | Mantener y analizar el sistema en conjunto |  |
| Administrador de seguridad |  | Analizar el tráfico de entrada y salida en la base de datos |  |

4.2 Convenciones de nomenclatura

| **Tipo** | **Línea Directriz** |
| --- | --- |
| Estilo | Usar caracteres en minúscula |
| Estilo | Separar palabras y prefijos con guiones bajos. Nunca usar espacios |
| Estilo | Evitar el uso de números |
| Nombre de Tabla | Escoger nombres cortos y no ambiguos. No más de dos palabras |
| Nombre de Tabla | Usar nombres en singular. Nunca en plural |
| Nombre de Tabla | Evitar acrónimos o nombres abreviados |
| Nombre de tabla | Concatenar los nombres de dos tablas que están enlazadas para una tabla enlazada |

4.3 Identificación de la base de datos.

| **Elemento** | **Nombre del elemento** | **Significado** |
| --- | --- | --- |
| db\_word | Base de datos de palabras | La base de datos donde se ubican las palabras según los distintos criterios. |
| fs\_word | Identificador único para los datos | Cada palabras es en sí misma una palabra clave con la que se identifica. |
| fs\_grammar\_clasification | Clasificación gramática de la palabra | Clasificación de la palabra determinada según su función gramatical. |
| fs\_common\_use | Clasificación de la palabra según su uso frecuente | Da un puesto a la palabra según su frecuencia de uso en el idioma. |
| fs\_semantic\_clasification | Clasificación semántica de la palabra | Se etiqueta a la palabra según la función semántica que cumple la mayoría de veces. |

**5 Acceso a los datos**

5.1 Definiciones de roles

| Nombre de rol | Objetivo |
| --- | --- |
| db\_owner | La persona de este rol puede realizar todas las actividades de configuración y mantenimiento en la base de datos y también pueden quitar la base de datos |
| db\_securityadmin | Los miembros del rol fijo de base de datos db\_securityadmin pueden modificar la pertenencia a roles únicamente para roles personalizados y administrar permisos |
| db\_accessadmin | Los miembros del rol fijo de base de datos db\_accessadmin pueden agregar o quitar el acceso a la base de datos para inicios de sesión de Windows. |
| db\_backupoperator | La persona de este rol puede crear copias de seguridad de la base de datos |
| db\_datawriter | La persona de este rol puede agregar, eliminar en la tabla de usuarios. |
| db\_datareader | Los miembros de la base de datos pueden leer todos los datos de todas las tablas. |

5.2 Usuarios

| Username | Objetivo |
| --- | --- |
| Modesto Calixto, Keler | db\_owner |
| Ascue Orosco, Carlos | db\_securityadmin |
| Apaza Alva, Rodrigo | db\_accessadmin |
| Triveno Ruffner Daniel | db\_backupoperator |
| Huaripata Huaccha, Héctor | db\_datawriter |
| Vara Vela, Neil | db\_datareader |
| Bejar Bonifacio Manuel | db\_datareader |

5.3 Plantillas de acceso a tablas

| Users |  |
| --- | --- |
| Horario |  |
| Recepcionista |  |
| Mensaje |  |

**6 Consideraciones de Implementación**

6.1 objetos grandes

6.2 colas

6.3 Particionamiento

**7.1 Diseño de seguridad**

7.2 Disponibilidad

7.3 escalabilidad

7.4 Rendimiento

7.5 Manejo de errores

7.6 Copias de seguridad y restauración

7.7 Archivado

