**“Prototipo de asistente corrector gramatical y ortográfico”**

***Trabajo Terminal No. — — — — -— — —***

Alumnos: Aguirre Hernández Leonardo Miguel, Álvarez Hernández Zeth.

Directores: Suarez Castañón Miguel Santiago

e-mail: leonardomaguirreh@gmail.com, zethalvarezh@hotmail.com

**Resumen** - Desarrollo de un prototipo de software, basado en una arquitectura orientada a servicios, en particular la denominada software como servicio o SaaS por sus siglas en inglés. Esta arquitectura es atractiva porque combina el poder de los sistemas orientados a servicios, con el poder y disponibilidad de Internet. El prototipo ha sido definido como: “Prototipo de asistente corrector gramatical y ortográfico en la redacción de protocolos correctos”, el cual para fines de presentación será nombrado: Prototipo de asistente corrector gramatical y ortográfico.

**Palabras clave** - software como un servicio, corrección ortográfica y gramatical automatizada, machine learning, procesamiento de lenguaje natural.

**1. Introducción**

Actualmente, el mundo está constantemente en comunicación gracias a la Internet y a la cantidad de herramientas que esta posee. Desde hace algunos unos años, para las personas es más práctico comunicarse mediante mensajes (a través de alguna red social), que realizar una llamada. Desafortunadamente, la gran mayoría es descuidada, o no le da importancia a la calidad de la redacción ni a la ortografía. Además, se han desarrollado una serie de “modismos” que demeritan la calidad de los mensajes. Sin embargo, lo común de estas prácticas, no es una justificación para que no nos preocupemos por redactar correctamente. No solo en el ámbito profesional, también el personal.

En México, de acuerdo con una investigación de Leobardo Hernández Audelo, maestro en Ciencias Computacionales por la UNAM, más del 50% de la población está conectada a Internet y son usuarios de perfiles de alguna red social [1].

Tras conocer este dato, nos dimos a la tarea de investigar si el uso de redes sociales influye en la redacción de los estudiantes de la ESCOM (Escuela Superior de Cómputo). Para esto, se analizaron las conversaciones de 20 alumnos, con la finalidad de identificar si su redacción era correcta. Una vez analizadas las conversaciones, llegamos a la conclusión de que 14 de 20 tenían al menos dos errores gramaticales u ortográficos.

Esto nos indica que más de la mitad de la población de la ESCOM está en constante comunicación por medio de mensajes, pero también, que hay una alta probabilidad de que no escriban de forma correcta.

A partir de lo expuesto, decidimos desarrollar un prototipo de software, que actúe como asistente corrector para la redacción de protocolos gramatical y ortográficamente correctos. Para esto, utilizaremos técnicas de inteligencia artificial, en particular algoritmos de machine learning enfocados al procesamiento y análisis de textos, mejor conocidos como algoritmos de procesamiento de lenguaje natural (PLN), los cuales nos permitirán detectar en su mayoría los errores ortográficos y gramaticales que se presenten [2].

El software estará disponible como un servicio en la nube, ya que, en su mayoría, la comunicación que se realiza actualmente a través de medios digitales es de forma escrita [3]. Disponer de un asistente en tiempo real, que ayude a pulir la escritura además de ayudarnos a escribir correctamente, nos ayudará a conocer mejor las reglas gramaticales y ortográficas, porque el asistente al aconsejarnos, justificará la sugerencia señalando la o las reglas que estaríamos infringiendo o las que el asistente estaría aplicando.

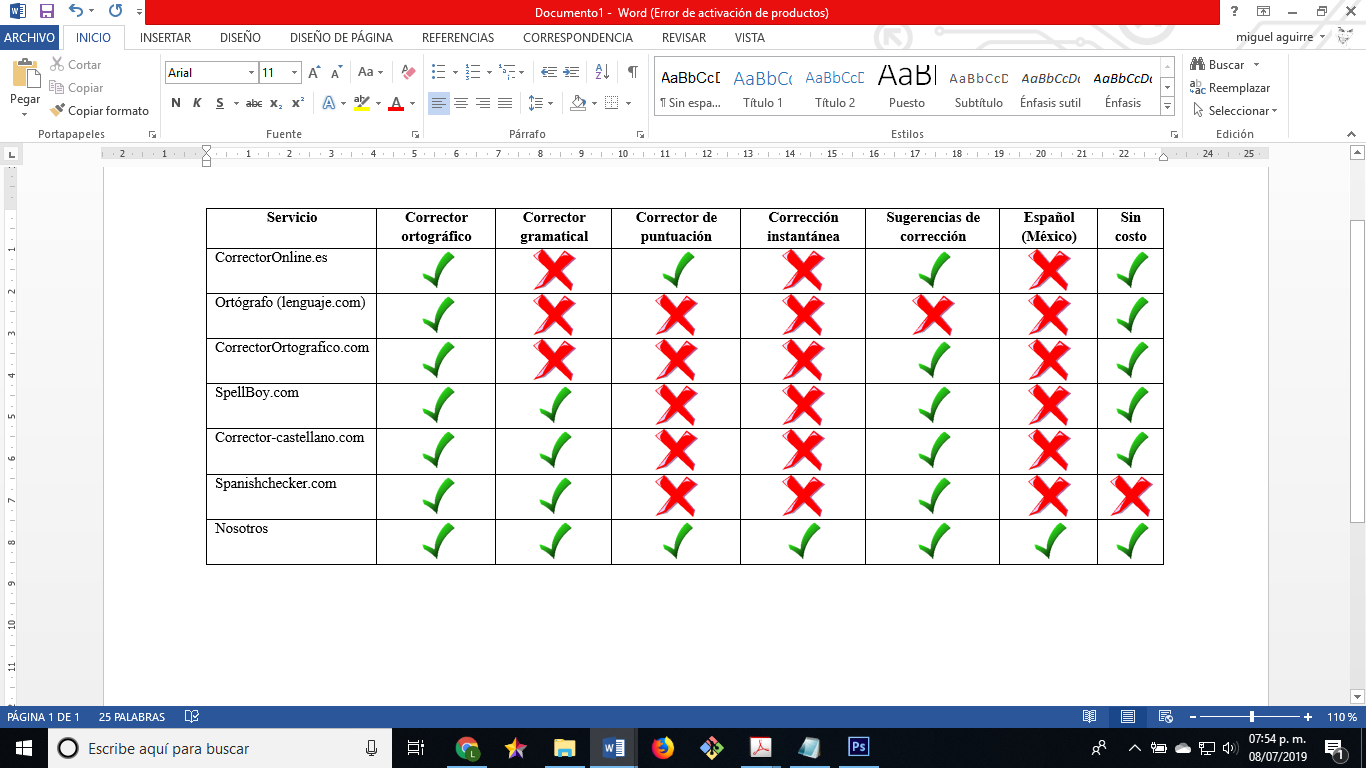
Se espera que el prototipo, objeto de este trabajo, pueda realizar correcciones ortográficas y gramaticales en la redacción de protocolos, y proporcionar sugerencias de acuerdo al tipo de error presentado. Con base en esto, el usuario analizará las posibles soluciones y recomendaciones que puede utilizar para mejorar sus escritos.

**Sistemas similares**

Una revisión exhaustiva de todas las aplicaciones existentes va más allá de los alcances de este trabajo terminal, por lo que hemos citado las que consideramos más importantes y que están íntimamente relacionadas con la propuesta objeto de este documento:

* **Correctoronline.es**: herramienta para detectar errores ortográficos, y algunos otros básicos de gramática y estilo, en textos en español.
* **Ortógrafo**: corrector ortográfico que cuenta con el léxico más completo de este idioma (español) formado por más de cinco millones de palabras. Incluye términos técnicos de diferentes áreas como agricultura, biología y comercio, entre otras. Reconoce todas las formas conjugadas, así como diminutivos aumentativos, sufijos y prefijos más utilizados.
* **CorrectorOrtografico.com**: corrector ortográfico de más de 11 idiomas que permite la corrección de textos con hasta 1000 caracteres.
* **SpellBoy.com:** servicio que encuentra muchos errores que un simple corrector ortográfico no puede detectar. Además de los errores ortográficos normales, también detecta fallos gramaticales y estilísticos.
* **Corrector-Castellano.com:** corrector ortográfico y gramatical en español, ideal para escribir correctamente tus textos sin faltas de ortografía.
* **Spanishchecker.com**: servicio que detecta los errores cometidos por los estudiantes de español y ayuda a mejorar sus habilidades de escritura.

En la tabla 1, se muestra la comparativa de las aplicaciones que se describen arriba, contra el prototipo que desarrollaremos en nuestro trabajo terminal.



**Tabla 1**. Comparativa de aplicaciones.

**2. Objetivo**

**Objetivo general**

* Realizar una herramienta que ayude a los estudiantes que estén cursando o hayan cursado el nivel superior de estudios a redactar correctamente protocolos en el idioma español (español de México).

**Objetivos particulares**

* Determinar qué métodos de procesamiento de lenguaje natural se deben implementar en el software.
* Realizar un algoritmo de corrección ortográfica.
* Realizar un algoritmo de corrección gramatical.
* Desplegar el asistente corrector en una plataforma orientada en SaaS utilizando una arquitectura SOA.

**3. Justificación**

De acuerdo con un reportaje realizado por la BBC Future el 6 de febrero de 2018:

En 2011, Daria Kuss y Mark Griffiths, de la Universidad Nottingham Trent, en Reino Unido, analizaron 43 estudios previos sobre el tema y llegaron a la conclusión de que la adicción a las redes sociales es un problema mental que "podría" requerir tratamiento profesional.

Se encontraron con que un uso excesivo iba asociado a los problemas en las relaciones personales, un peor rendimiento académico y una menor participación en comunidades reales y no virtuales [4].

El uso excesivo de las redes sociales en México abre paso a generar malos hábitos al redactar mensajes, debido a que, al redactar dentro de las mismas, no se tiene la misma formalidad que al redactar un documento profesional, permitiendo que esto no quede únicamente dentro de la red social, sino que pase a afectar en otros ámbitos, viéndose severamente afectado el laboral.

Es por eso que, el prototipo de este sistema está enfocado en desarrollar un asistente inteligente, que brinde ayuda corrigiendo determinados errores ortográficos y gramaticales en la redacción de protocolos. Con esto, además de mejorar la calidad de los documentos, se espera que proporcione sugerencias para que el usuario reciba una retroalimentación de sus errores.

Esto se decidió porque existe una gran pérdida en la calidad del lenguaje y una reducción en la capacidad de los estudiantes para redactar, efecto de tener una forma libre de escribir por Internet, aunque hay quienes aseguran que el problema no son los medios digitales, sí no la misma educación de los jóvenes [5].

El prototipo, pretende ser una herramienta de apoyo que cubra la necesidad de corregir errores en la redacción de protocolos en el idioma español de México. Ya que, por lo visto en el estado del arte, no existe una herramienta que cumpla adecuadamente con asistencia en esta lengua.

Implementaremos algoritmos de PLN que lo harán independiente, esto quiere decir que, de alguna manera, el sistema aprenderá con base en ejemplos reales a distinguir y corregir errores ortográficos y gramaticales, sin la necesidad de un administrador que ingrese manualmente las reglas que rigen el idioma español.

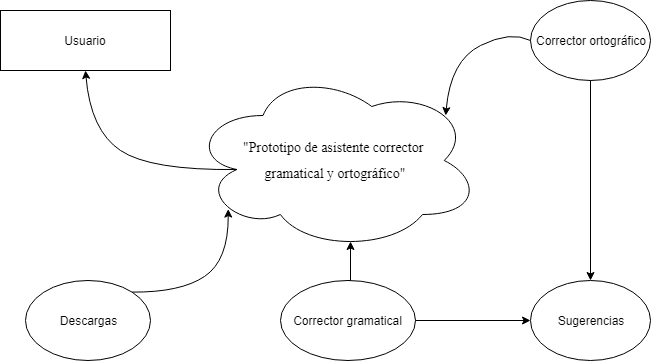
Para el desarrollo, usaremos una metodología ágil, que nos permita trabajar más rápido sobre cambios y errores que surjan y reducir la documentación excesiva [6].

**4. Resultados esperados**

De acuerdo con la investigación previamente realizada, el prototipo tendrá los siguientes módulos:

* Módulo de corrección ortográfica: hará la detección y corrección de las palabras que estén mal escritas.
* Módulo de corrección gramatical: identificará la coherencia entre las palabras empleadas en una oración y detectará posibles errores.
* Módulo de sugerencias: de acuerdo al tipo de error que se presente, mostrará las posibles soluciones, así como una breve justificación para cada solución.
* Módulo de descargas: permitirá descargar el contenido del cuadro de texto en un formato .txt o .pdf, según lo indique el usuario.

En la figura 1 se muestra como estarán conectados los módulos descritos arriba para poder interactuar de forma adecuada y brindar buenos resultados utilizando la arquitectura SOA.

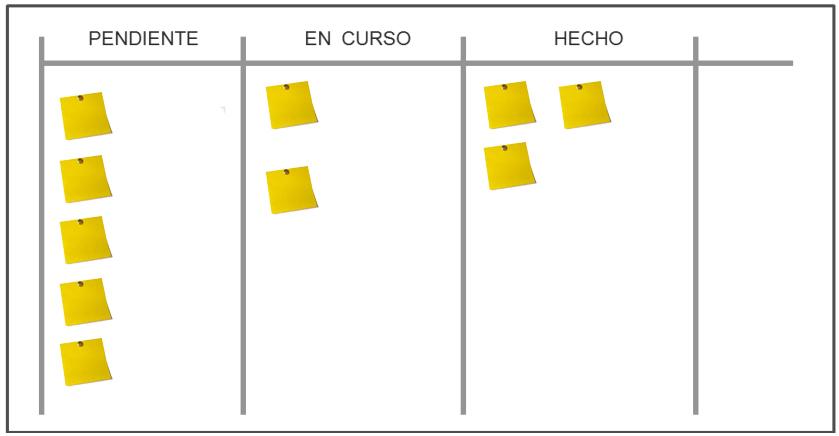


**Figura 1.** Diagrama de contexto.

**5. Metodología**

Kanban se basan en el desarrollo incremental, dividiendo el trabajo en partes. Normalmente cada una de esas partes se escribe en una nota y se pega en un tablero. Las notas suelen tener información variada, si bien, aparte de la descripción, debieran tener la estimación de la duración de la tarea. El tablero tiene tantas columnas como estados por los que puede pasar la tarea [7].

A continuación, en la figura 2, se muestra un ejemplo del tablero de la metodología Kanban.



**Figura 2**. Ejemplo del tablero de la metodología Kanban.

Algunas de las ventajas al implementar Kanban son:

* Facilidad de entendimiento.
* Visibilidad de la información a todos los miembros involucrados.
* Facilidad de integración con metodologías ágiles (Scrumban).
* El más adecuado para proyectos que se encuentran en mejora continua.
* Acepta el ingreso de cambios a último momento con facilidad.

**Arquitectura:**

**SOA, Service oriented architecture (Arquitectura orientada a servicios)**

SOA actualmente, significa que los componentes de una aplicación actúan como servicios interoperables y pueden ser usados de manera independiente y combinados con otros servicios. La implementación en contraste es considerada "software silo" (cualquier sistema de gestión que no pueda operar con ningún otro sistema), el cual raramente tiene interfaces de programación de aplicaciones (API) externalizables a componentes internos.

Si te equivocas respecto a lo que el cliente quiere en realidad, el costo es mucho menor con SOA que con “software silo” para recuperarse de los errores e intentar algo más o para producir una variante similar, pero no idéntica para complacer a un subconjunto de usuarios.

**SaaS, Software as a Service (software como servicio)**

El poder de SOA combinado con el poder de la Internet da espacio a un caso especial de SOA con su propio nombre: Software as a Service (SaaS). Ofrece software y datos como un servicio a través de la Internet, por lo general, a través de un programa como un navegador que se ejecuta en dispositivos cliente locales.

Las ventajas para el cliente y para el desarrollador de software son grandes:

1. Los clientes no necesitan instalar la aplicación, no tienen que preocuparse si su hardware es de la marca correcta o lo suficientemente rápido, ni si tienen la versión correcta del sistema operativo.
2. Los datos asociados con el servicio generalmente se guardan con el servicio, por lo que los clientes no deben preocuparse por hacer una copia de seguridad, perderlo debido a un mal funcionamiento del hardware local o incluso perder todo el dispositivo, como un teléfono o tableta.
3. Cuando un grupo de usuarios desea interactuar colectivamente con los mismos datos, SaaS es un vehículo natural.
4. Cuando los datos son grandes y / o se actualizan con frecuencia, puede tener más sentido centralizar los datos y ofrecer acceso remoto a través de SaaS.
5. Solo una copia del software del servidor se ejecuta en un entorno de sistema operativo y hardware uniforme y estrechamente controlado, seleccionado por el desarrollador, lo que evita los problemas de compatibilidad de la distribución de binarios que deben ejecutarse en computadoras y sistemas operativos de gran envergadura.
6. Los desarrolladores pueden probar nuevas versiones de la aplicación en una pequeña fracción de los clientes reales temporalmente sin molestar a la mayoría de los clientes.
7. Los desarrolladores no necesitan molestar a los usuarios con solicitudes de permiso para actualizar sus aplicaciones [8].

**Patrón de diseño de software:**

**Health Endpoint Monitoring (monitoreo de puntos finales de salud)**

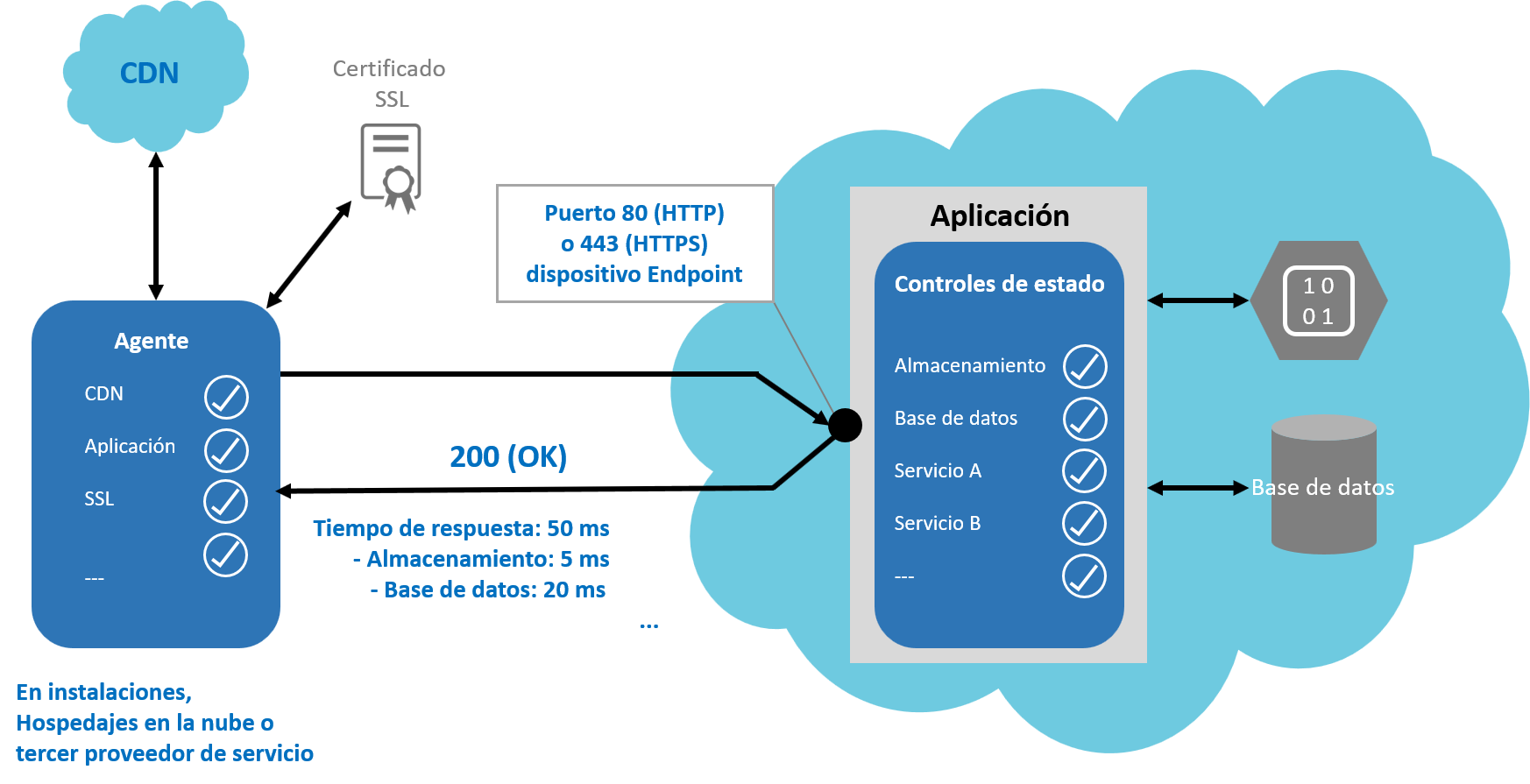
Prototipo de asistente corrector gramatical y ortográfico es un prototipo de software que, por su funcionalidad, debe trabajar de tal forma que el usuario no deba cargar una y otra vez la misma ventana. Esto se vuelve una necesidad, que forzosamente requiere de una conexión asíncrona.

Por tal motivo, “Health Endpoint Monitoring” resulta el patrón de diseño adecuado para el desarrollo del prototipo, ya que implementa verificaciones funcionales en una aplicación a las que puedan acceder las herramientas externas a través de puntos finales expuestos a intervalos regulares. Esto puede ayudar a verificar que las aplicaciones y los servicios estén funcionando correctamente [9].

Una verificación de monitoreo de salud típicamente combina dos factores:

* Las verificaciones (si las hay) realizadas por la aplicación o el servicio en respuesta a la solicitud al punto final de verificación de estado.
* Análisis de los resultados por la herramienta o marco que realiza la verificación de estado.

El código de respuesta indica el estado de la aplicación y, opcionalmente, cualquier componente o servicio que utiliza. La verificación de la latencia o el tiempo de respuesta se realiza mediante la herramienta de monitoreo o marco. De acuerdo con el sitio web de Microsoft, la figura 3 proporciona una visión general del patrón.



**Figura 3**. Funcionamiento del patrón de diseño “Health Endpoint Monitoring”.

Otras comprobaciones que podrían ser realizadas por el código de monitoreo de salud en la aplicación incluyen:

* Verificación de almacenamiento en la nube o una base de datos para disponibilidad y tiempo de respuesta.
* Verificar otros recursos o servicios ubicados en la aplicación, o ubicados en otro lugar, pero utilizados por la aplicación.

**6. Cronograma**

Ver anexos 1, 2.

**7. Referencias**

[1] "México, cuarto lugar a nivel mundial en uso de redes sociales", *Excélsior*, 2018. [En línea].

Disponible en: https://www.excelsior.com.mx/hacker/2018/01/18/1214650?fbclid=IwAR2b4vqV5jS1vF\_eG7Fs6c5MwqyJ7kzBqVKR5ZsOHO3Ev3TKE8niA\_t5v6I.

[Accedido: 08 - Julio - 2019]

[2] A. Kulkarni and A. Shivananda, Natural Language Processing Recipes. Berkeley, CA: Apress, 2019, p. 2.

[3] D. Crovi Druetta, "¿ES INTERNET UN MEDIO DE COMUNICACIÓN?", Revista Digital Universitaria, vol. 7, pp. 4-5, 2006. [Accedido: 08 - Julio - 2019]

[4] ¿Realmente las redes sociales son perjudiciales para tu salud?", BBC News Mundo, 2019. [En línea]. Disponible en: https://www.bbc.com/mundo/vert-fut-42628812. [Accedido: 08 - Julio - 2019]

[5] Boza, M. (2016). LA ESCRITURA EN LOS ENTORNOS DIGITALES. educare, 20, pp.155-156.

[6] A. Fox, D. Patterson and S. Joseph, Engineering software as a service. 2a Edición, Strawberry Canyon LLC., 2013, pp. 35 – 45

[7] Executive master in project management, Universidad de Alcalá, "Gestión ágil de proyectos con kanban", 2014 [En línea]. Disponible en: <http://www.uv-mdap.com/programa-desarrollado/bloque-iv-metodologias-agiles/gestion-agil-de-proyectos-con-kanban/> [Accedido: 09 - Julio - 2019]

[8] A. Fox, D. Patterson and S. Joseph, Engineering software as a service. 2a Edición, Strawberry Canyon LLC., 2013, pp. 46 – 53

[9] Health Endpoint Monitoring pattern - Cloud Design Patterns", Docs.microsoft.com, 2019. [En línea]. Disponible en: https://docs.microsoft.com/es-es/azure/architecture/patterns/health-endpoint-monitoring. [Accedido: 08 - Julio - 2019]

**CARÁCTER: Confidencial**

**FUNDAMENTO LEGAL: Art. 3, fracc. II, Art. 18, fracc. II**

**y Art. 21, lineamiento 32, fracc. XVII de la L.F.T.A.I.P.G.**

**PARTES CONFIDENCIALES: No. de boleta y Teléfono.**

**8. Alumnos y directores**

Leonardo Miguel Aguirre Hernández. - Alumno de la carrera de Ing. En Sistemas Computacionales en ESCOM, Especialidad: Sistemas, Boleta: 2014630008, Tel.5531692807, email: leonardomaguirreh@gmail.com

Firma: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Zeth Álvarez Hernández. - Alumno de la carrera de Ing. En Sistemas Computacionales en ESCOM, Especialidad: Sistemas, Boleta: 2014310030, Tel.: 5547131440, email: zethalvarezh@hotmail.com.

Firma: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Miguel Santiago Suárez Castañón. - Ingeniero en cibernética y ciencias de la computación por la universidad LaSalle, A.C. Obtuvo el grado de maestro en ciencias de la computación en la Universidad Nacional Autónoma de México y el grado de doctor en ciencias de la computación en el Instituto Politécnico Nacional. En 2005 obtuvo el premio a la mejor tesis doctoral. Es miembro del Sistema Nacional de Investigadores nivel I desde el año 2007. Es co-autor de más de 30 artículos científicos publicados en revistas contenidas dentro del Journal Citation Report.

Firma: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Anexo 1

Cronograma

Nombre: Leonardo Miguel Aguirre Hernández

Nombre del TT: Prototipo de asistente corrector gramatical y ortográfico

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Actividad** | **Julio** | **Agosto** | **Septiembre** | **Octubre** | **Noviembre** | **Diciembre** |
| Estructuración de protocolo |  |  |  |  |  |  |
| Análisis y diseño del sistema |  |  |  |  |  |  |
| Especificación de casos de uso |  |  |  |  |  |  |
| Definir tecnologías a utilizar |  |  |  |  |  |  |
| Generación de código (corrector ortográfico) |  |  |  |  |  |  |
| Despliegue del sistema en plataforma compatible con SaaS |  |  |  |  |  |  |
| Pruebas |  |  |  |  |  |  |
| Generación de manuales técnico y de usuario. |  |  |  |  |  |  |
| Evaluación de TTR |  |  |  |  |  |  |

Anexo 2

Cronograma

Nombre: Zeth Álvarez Hernández

Nombre del TT: Prototipo de asistente corrector gramatical y ortográfico

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Actividad** | **Julio** | **Agosto** | **Septiembre** | **Octubre** | **Noviembre** | **Diciembre** |
| Estructuración de protocolo |  |  |  |  |  |  |
| Análisis y diseño del sistema |  |  |  |  |  |  |
| Especificación de casos de uso |  |  |  |  |  |  |
| Definir tecnologías a utilizar |  |  |  |  |  |  |
| Generación de código (corrector gramatical) |  |  |  |  |  |  |
| Generación de código (Modulo de descargas) |  |  |  |  |  |  |
| Pruebas |  |  |  |  |  |  |
| Generación de manuales técnico y de usuario. |  |  |  |  |  |  |
| Evaluación de TTR |  |  |  |  |  |  |