

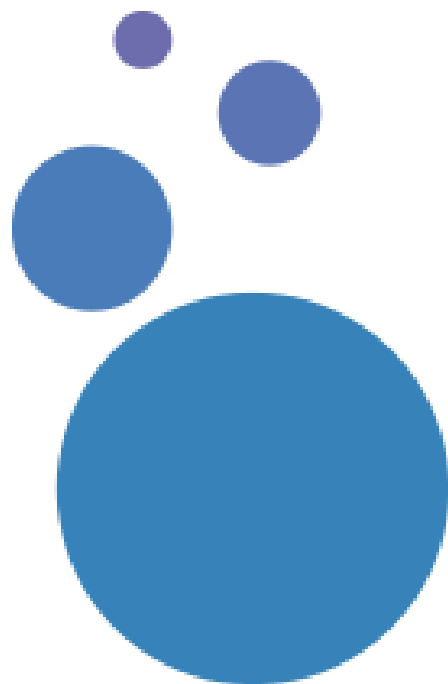


# **NORMA: La abogada de las víctimas**

**Memoria técnica**

**INSTITUTO MEXICANO  
DE LA COMPETITIVIDAD**

**Septiembre 2018**



## 1. ÍNDICE

1.	RESUMEN EJECUTIVO.....	1
2.	OBJETIVO .....	2
3.	PRINCIPALES RETOS .....	2
4.	COMPONENTES DE LA SOLUCIÓN.....	6
4.1.	GATEWAYS DE COMUNICACIÓN.....	7
4.2.	ANALIZADOR DE LENGUAJE NATURAL.....	8
4.3.	SISTEMA ADMINISTRATIVO.....	9
4.4.	BASES DE DATOS.....	11
5.	TECNOLOGÍAS Y AMBIENTES DE EJECUCIÓN.....	12
5.1.	GATEWAYS DE COMUNICACIÓN.....	12
5.2.	ANALIZADOR DE LENGUAJE NATURAL.....	13
5.3.	SISTEMA ADMINISTRATIVO.....	14
5.4.	BASES DE DATOS.....	16

## 1. RESUMEN EJECUTIVO

El presente documento contiene detalles de la solución, que OPIT implementó en Instituto Mexicano para la Competitividad (IMCO), relacionada con una plataforma computacional que permite otorgar asesoría jurídica de manera automatizada a través de un modelo de conocimiento integrado.

De acuerdo con el INEGI, en 2015 se cometieron 29.3 millones de delitos en México, de los cuales el 93.7% no se denunció (9 de cada 10). Para subsanar la problemática mencionada, el Instituto Mexicano para la Competitividad (IMCO) participó en el desafío Google 2017 en México, donde resultó uno de los tres ganadores con el proyecto “Norma, la abogada de las víctimas” (NORMA).

El objetivo de NORMA es (i) apoyar a las víctimas para incentivar las denuncias y (ii) mejorar la atención a los denunciantes, mediante una herramienta de asesoría digital que guía paso a paso, a través de mensajes SMS y de un chat, a las víctimas del delito en el proceso de denuncia. NORMA es una solución en tiempo real, que atiende de forma anónima y sin costo para los usuarios.

Asimismo, se pretende que NORMA sea una fuente de datos para mejorar las políticas públicas en materia de procuración de justicia.

La plataforma contará con cuatro pilares normativos:

- Gateways de mensajería. Son los proveedores que permiten el intercambio de comunicación con los usuarios finales.
- Analizador de lenguaje natural. Es el componente que permite interpretar los mensajes recibidos por los gateways de mensajería.
- Gestor de modelo de conocimiento. Permitirá definir toda la lógica jurídica aplicable en el sistema. Asimismo, permitirá administrar gráficamente el árbol de decisiones relevante.
- Base de datos. Permitirá guardar la información del modelo de conocimiento, así como de las interacciones entre los usuarios y NORMA.

## 2. OBJETIVO

El objetivo es desarrollar una plataforma que permita (i) interactuar con usuarios a través de mensajes de texto SMS; (ii) analizar los mensajes recibidos a través de un componente analizar de lenguaje natural; (iii) integrar los agentes encontrados en los mensajes de texto con un modelo de conocimiento definido a priori; y (iv) analizar la interacción del sistema con los usuarios para el mejoramiento de la inteligencia del sistema.

## 3. PRINCIPALES RETOS

FUNCIONALIDAD	DESCRIPCIÓN
<b>GENERALIDAD DE LA SOLUCIÓN</b>	<ul style="list-style-type: none"><li>• Se requiere que las tecnologías utilizadas para el desarrollo de la infraestructura permita la portabilidad de la solución a lo largo de diferentes infraestructuras.<ul style="list-style-type: none"><li>○ La solución evita dependencias innecesarias de proveedores de infraestructura.</li></ul></li><li>• Para evitar problemas en la comunicación de componentes de la plataforma, se sugiere un mecanismo genérico.<ul style="list-style-type: none"><li>○ Para aislar a los componentes del ambiente de ejecución de otros componentes, se propone que la comunicación entre éstos sea a través de APIs</li><li>○ Para permitir la interconexión dinámica de los componentes, se propone la utilización del protocolo Restful en la implementación de las APIs.</li></ul></li><li>• Para asegurar la generalidad de la solución, se propone que la plataforma contenga algunos componentes administrados “in-house”, mientras que otros componentes sean adquiridos como servicio.</li></ul>

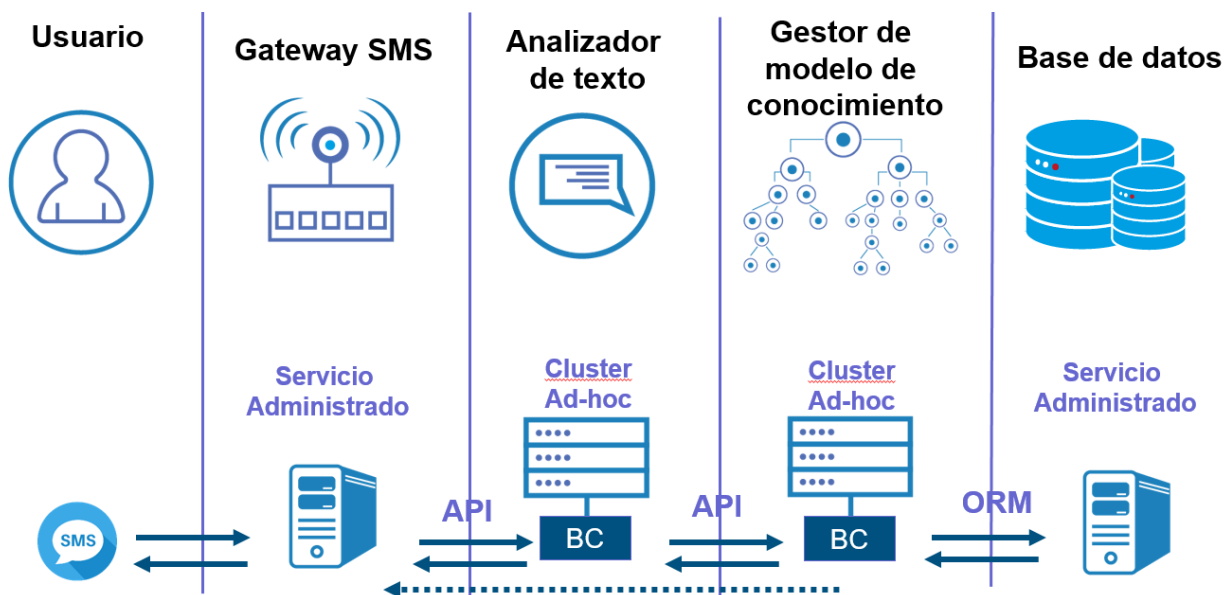
FUNCIONALIDAD	DESCRIPCIÓN
<b>AGNOSTICIDAD CON GATEWAYS DE COMUNICACIÓN.</b>	<ul style="list-style-type: none"> <li>• La plataforma garantiza total independencia con respecto a los proveedores de mensajería SMS. <ul style="list-style-type: none"> <li>○ La independencia del proveedor permitirá la flexibilidad necesaria, dado que se requiere un control estricto de gastos.</li> </ul> </li> <li>• Para permitir la comunicación entre la plataforma NORMA y los usuarios finales, se requiere que los proveedor de mensajería SMS tengas disponible la solución “two-way messaging”. <ul style="list-style-type: none"> <li>○ La solución “two-way messaing” permite controlar un punto de acceso virtual para el envío y recepción de mensajes SMS.</li> </ul> </li> <li>• Dado que se espera que el número promedio de mensajes (entrantes y salientes) sea relevante, se requiere que el proveedor cuente con un “throughput” mínimo aceptable. <ul style="list-style-type: none"> <li>○ Se entiende como “throughput” a la cantidad máxima de transmisión de mensajes en un lapso de tiempo.</li> </ul> </li> </ul>
<b>GENERACIÓN DE MODELO DE CONOCIMIENTO.</b>	<ul style="list-style-type: none"> <li>• Definición de principales casos de uso.</li> <li>• Mapear el modelo de conocimiento a un modelo inteligible computacionalmente.</li> </ul>
<b>CONTROL DE FLUJO DE INFORMACIÓN PARA ANÁLISIS FORENSE.</b>	<ul style="list-style-type: none"> <li>• Guardar las interacciones con el usuario en bases de datos secundarias.</li> <li>• Análisis de las interacciones con los usuarios para mejorar el sistema de reconocimiento de texto de manera iterativa.</li> </ul>

FUNCIONALIDAD	DESCRIPCIÓN
ESCALABILIDAD	<ul style="list-style-type: none"><li>• Para asegurar la escalabilidad de la plataforma NORMA, se requiere que los componentes administrados “in-house” sean replicables de forma nativa en diferentes infraestructuras.<ul style="list-style-type: none"><li>○ Para asegurar la replicabilidad, la solución debe estar basada en algoritmos asíncronos.</li></ul></li><li>• Para asegurar la escalabilidad de la plataforma NORMA, se requiere que el throughput de todos los componentes sean elásticos.<ul style="list-style-type: none"><li>○ En el caso de componentes operados como servicios de terceros, éstos deben contar con políticas definidas de throughput garantizado.</li><li>○ En el caso de los componentes administrados “in-house”, éstos deben ser modificables en sus partes críticas (a fin de su optimización).</li></ul></li><li>• Se debe considerar la necesidad de:<ul style="list-style-type: none"><li>○ Administración de servidores.</li><li>○ Administración de balanceadores de carga.</li><li>○ Administración de repositorios de archivos.</li><li>○ Administración de servicios DNS.</li></ul></li></ul>

FUNCIONALIDAD	DESCRIPCIÓN
<p><b>MINIMIZACIÓN DE COSTOS</b></p>	<ul style="list-style-type: none"> <li>• Costos de infraestructura. <ul style="list-style-type: none"> <li>○ Instancias de servidores alojados en la nube.</li> <li>○ Instancias de balanceadores de carga.</li> <li>○ Instancias de servicios adicionales de proveedores de infraestructura.</li> <li>○ Costos relacionados con los nombres de dominio utilizados.</li> </ul> </li> <li>• Costos de manejo de SMS. <ul style="list-style-type: none"> <li>○ Costo por arrendamiento de número telefónico desde el cual se enviarán y recibirán mensajes de terceros.</li> <li>○ Costo unitario por envío de mensajes SMS a usuarios finales.</li> </ul> </li> <li>• Costo de análisis de texto en mensajes (lenguaje natural). <ul style="list-style-type: none"> <li>○ Costo unitario por cadena de texto analizada.</li> </ul> </li> <li>• Costo de administración y mantenimiento de modelo de conocimiento. <ul style="list-style-type: none"> <li>○ Costo del personal que dará mantenimiento a los componentes de la plataforma que son administrados “in-house”</li> <li>○ Costo del personal que analizará las interacciones de la plataforma con los usuarios finales a fin de mejorar el modelo del analizador de lenguaje natural.</li> <li>○ Costo del personal que analizará y mejorará el modelo de conocimiento.</li> </ul> </li> </ul>

## 4. COMPONENTES DE LA SOLUCIÓN

Para el correcto funcionamiento de NORMA, la implementación utiliza una arquitectura basada en componentes intercambiables. En el siguiente diagrama se muestran los componentes que propuestos en la solución.



El flujo de actividad esperado en el funcionamiento de NORMA se describe a continuación.

1. El usuario envía un mensaje SMS al número contratado o inicia una conversación en el chat.
  - a. Comunicación SMS: el Gateway de mensajería SMS recibe el mensaje y lo reenvía un webhook del modelo de conocimiento.
  - b. Comunicación por chat: la tecnología asíncrona manda una petición con la información relvante al webhook del analizador del modelo de conocimiento.
2. El gestor del modelo de conicmiento analiza la información y lo envía al analizador de lenguaje natural para realizar una clasificación gramatical.
3. El analizador de lenguaje natural procesa el texto y lo clasifica gramaticalmente.



4. El analizador de texto envía el mensaje procesado y clasificado a un webhook del gestor del modelo de conocimiento.
5. El gestor del modelo de conocimiento detecta el flujo de información a seguir mediante una interacción bidireccional con la base de datos.
6. El gestor del modelo de conocimiento guarda la información de la interacción con el usuario en la base de datos.
7. El gestor del modelo de conocimiento regresa un mensaje de texto al analizador de lenguaje natural.
8. El analizador de lenguaje natural forma la cadena final y la envía al Gateway de mensajería SMS.
9. El Gateway de mensajería SMS envía el mensaje SMS al usuario.

#### 4.1. GATEWAYS DE COMUNICACIÓN.

FUNCIONALIDAD	DESCRIPCIÓN
<b>CARACTERÍSTICAS DEL SERVICIO</b>	<ul style="list-style-type: none"><li>• Se debe adquirir un servicio “two-way service” para poder interactuar con los usuarios.</li><li>• El Gateway debe tener capacidad de conexión mediante API para:<ul style="list-style-type: none"><li>○ Consultar mensajes enviados.</li><li>○ Consultar mensajes recibidos.</li><li>○ Enviar mensajes de texto.</li></ul></li></ul>

FUNCIONALIDAD	DESCRIPCIÓN
<b>ALCANCE</b>	<ul style="list-style-type: none"><li>• Se integró el API del Gateway de mensajes SMS con la plataforma NORMA para realizar las siguientes funcionalidades.<ul style="list-style-type: none"><li>○ Consultar mensajes enviados.</li><li>○ Consultar mensajes recibidos.</li><li>○ Enviar mensajes de texto.</li></ul></li></ul>

## 4.2. ANALIZADOR DE LENGUAJE NATURAL.

Existen múltiples soluciones para analizar texto a fin de obtener su estructura gramatical (sujeto, objeto y acción).

FUNCIONALIDAD	DESCRIPCIÓN
<b>CARACTERÍSTICAS DEL SERVICIO</b>	<ul style="list-style-type: none"><li>• Se requiere que el componente disponga de un API para establecer comunicación bidireccional con el gateway SMS y el sistema gestor de modelo de conocimiento.</li><li>• Se requiere que el API contenga como mínimo los siguientes métodos.<ul style="list-style-type: none"><li>○ Recuperación de mensajes de texto analizados.</li><li>○ Recepción de mensajes de texto para análisis.</li></ul></li><li>• Ejecución de análisis de texto en mensajes recibidos.</li></ul>

FUNCIONALIDAD	DESCRIPCIÓN
<p><b>ALCANCE</b></p>	<ul style="list-style-type: none"> <li>• Se integró el API del analizador de lenguaje natural con el Gateway de mensajes SMS.               <ul style="list-style-type: none"> <li>○ Ejecución de análisis de lenguaje en textos recibidos en el Gateway.</li> <li>○ Envío de mensajes SMS.</li> </ul> </li> <li>• Se integró el API del analizador de lenguaje natural con el Gestor del modelo de conocimiento.</li> <li>• Detección del siguiente mensaje de texto a enviar derivado de la ejecución del árbol de decisión en el contexto de la conversación.</li> </ul>

### 4.3. SISTEMA ADMINISTRATIVO.

FUNCIONALIDAD	DESCRIPCIÓN
<p><b>CARACTERÍSTICAS DEL SERVICIO</b></p>	<ul style="list-style-type: none"> <li>• Sistema administrado “in-house” con capacidad multi-usuario, escalable horizontalmente para atender picos de demanda.</li> <li>• Capacidad de conexión bidireccional con:               <ul style="list-style-type: none"> <li>○ Sistema de análisis de texto.</li> </ul> </li> <li>• Base de datos.</li> </ul>

<p><b>ALCANCE FUNCIONALIDAD DE INICIO DE SESIÓN</b></p>	<p>Inicio de sesión:</p> <ul style="list-style-type: none"> <li>• El usuario cuenta con usuario y contraseña únicos para acceder a los servicios que le sean permitidos.</li> <li>• Motor para el acceso, validación y asignación de roles por usuario.</li> </ul> <p>Administración de usuarios:</p> <ul style="list-style-type: none"> <li>• Se definió un esquema de permisos para diferentes tipos de usuarios dentro del sistema.</li> </ul> <p>Seguridad del sistema:</p> <ul style="list-style-type: none"> <li>• Se provee de diversas medidas de seguridad para garantizar la integridad de la información.</li> <li>• Las contraseñas de los usuarios de la plataforma de administración están encriptadas antes de guardarse en la base de datos, por lo que no se podrá ser accedida directamente.</li> </ul> <p>Conexión a la base de datos:</p> <ul style="list-style-type: none"> <li>• Se implementó una capa adicional para la conexión a la base de datos, con lo cual se abstraerá el acceso a ella mediante un sistema de permisos.</li> </ul> <p>Gestor de conocimiento:</p> <ul style="list-style-type: none"> <li>• Se definieron interfaces para definir árboles de decisión.</li> <li>• Se permiten definir parámetros a detectar en los textos recibidos: <ul style="list-style-type: none"> <li>○ Acción.</li> <li>○ Ubicación.</li> <li>○ Tipo de delito.</li> <li>○ Datos de contacto.</li> </ul> </li> <li>• Se permite recuperar conversaciones para análisis posterior.</li> </ul>
---	--

#### 4.4. BASES DE DATOS.

FUNCIONALIDAD	DESCRIPCIÓN
<b>CARACTERÍSTICAS DEL SERVICIO</b>	<ul style="list-style-type: none"><li>• Escalable horizontal y verticalmente.</li><li>• Compatible con SQL-ANSI.</li><li>• Módulos existentes para análisis de texto.</li></ul>

## 5. TECNOLOGÍAS Y AMBIENTES DE EJECUCIÓN.

En la siguiente imagen se describen las tecnologías y plataformas utilizadas durante la implementación de los diferentes módulos.



### 5.1. GATEWAYS DE COMUNICACIÓN.

FUNCIONALIDAD	DESCRIPCIÓN
TECNOLOGÍAS	<ul style="list-style-type: none"> <li>• Propietario.</li> <li>• El proveedor es Quiubas.</li> </ul>
AMBIENTE DE EJECUCIÓN	<ul style="list-style-type: none"> <li>• NA</li> </ul>
REQUERIMIENTOS	<ul style="list-style-type: none"> <li>• Se utiliza una librería en python del proveedor para la integración con el gestor del modelo de conocimiento.</li> </ul>

FUNCIONALIDAD	DESCRIPCIÓN
ESPECIFICACIONES	<p>La interacción con el Gateway SMS ocurre en dos momentos:</p> <ol style="list-style-type: none"> <li>1) Recepción de mensajes:               <ol style="list-style-type: none"> <li>a. La recepción se hace a través de un webhook que recibe la petición JSON del proveedor.</li> <li>b. El webhook se define en el panel de administración del proveedor.</li> </ol> </li> <li>2) Envío de mensajes:               <ol style="list-style-type: none"> <li>a. El envío se hace realizando peticiones JSON bajo demanda a un API del proveedor.</li> </ol> </li> </ol>

## 5.2. ANALIZADOR DE LENGUAJE NATURAL.

Existen múltiples soluciones para analizar texto a fin de obtener su estructura gramatical (sujeto, objeto y acción).

FUNCIONALIDAD	DESCRIPCIÓN
TECNOLOGÍAS	<ul style="list-style-type: none"> <li>• Propietario.</li> <li>• El proveedor es Google mediante su servicio DialogFlow.</li> </ul>
AMBIENTE DE EJECUCIÓN	<ul style="list-style-type: none"> <li>• NA</li> </ul>

FUNCIONALIDAD	DESCRIPCIÓN
REQUERIMIENTOS	<ul style="list-style-type: none"> <li>• Se requiere abrir una cuenta en Dialogflow si ésta será gratuita (quedará sujetas a políticas del proveedor) o en Google Cloud Platform si se pagará el servicio empresarial.</li> <li>• La interacción con los servicios de Dialogflow se realiza utilizando una librería de Python que el proveedor liberó al público.</li> <li>• La librería se encarga del proceso de autenticación y del envío de peticiones y recepción de respuestas.</li> </ul>
ESPECIFICACIONES	<ul style="list-style-type: none"> <li>• El gestor del modelo de conocimiento será el encargado de determinar cuando se requiere del análisis de lenguaje natural por parte de DialogFlow.</li> </ul>

### 5.3. SISTEMA ADMINISTRATIVO.

FUNCIONALIDAD	DESCRIPCIÓN
TECNOLOGÍAS	<ul style="list-style-type: none"> <li>• Lenguajes de programación: <ul style="list-style-type: none"> <li>○ Python</li> </ul> </li> <li>• Metalenguajes: <ul style="list-style-type: none"> <li>○ HTML y XML</li> </ul> </li> <li>• Frameworks: <ul style="list-style-type: none"> <li>○ ODOO</li> </ul> </li> </ul>



FUNCIONALIDAD	DESCRIPCIÓN
<b>AMBIENTE DE EJECUCIÓN</b>	<ul style="list-style-type: none"> <li>• Se utiliza la infraestructura SAS de Linode.</li> <li>• Se utiliza un sistema operativo Linux Ubuntu 16.10+</li> <li>• La versión utilizada de ODOO es Community 11</li> </ul>
<b>REQUERIMIENTOS</b>	<ul style="list-style-type: none"> <li>• Se tienen las siguientes dependencias directas:</li> </ul> <p>python-dateutil python-docutils python-feedparser python-jinja2 python-ldap python-libxslt1 python-lxml python-mako python-mock python-openid python-psycpg2 python-psutil python-pybabel python-pychart python-pydot python-pyparsing python-reportlab python-simplejson python-tz python-unittest2 python-vatnumber python-vobject python-webdav python-werkzeug python-xlwt python-yaml python-zsi poppler-utils python-pip python-pypdf python-passlib python-decorator gcc python-dev mc bzip python-setuptools python-markupsafe python-reportlab-accel python-zsi python-yaml python-argparse python-openssl python-egenix-mxdatetime python-usb python-serial lptools make python-pydot python-psutil python-paramiko poppler-utils python-pdftools antiword python-requests python-xlswriter python-suds python-psycogreen python-ofxparse python-gevent</p> <ul style="list-style-type: none"> <li>• Se requiere una instancia de PostgreSQL accesible por ODOO.</li> </ul>
<b>ESPECIFICACIONES</b>	<ul style="list-style-type: none"> <li>• Los objetos, la lógica y la comunicación se realiza mediante scripts en Python, que son interpretados por el motor del framework ODOO.</li> <li>• Las vistas y layouts se definen mediante archivos XML que son interpretados por el motor del framework ODOO.</li> </ul>

## 5.4. BASES DE DATOS.

FUNCIONALIDAD	DESCRIPCIÓN
TECNOLOGÍAS	<ul style="list-style-type: none"><li>• PostgreSQL</li></ul>
AMBIENTE DE EJECUCIÓN	<ul style="list-style-type: none"><li>• Se utiliza la infraestructura SAS de Linode.</li><li>• Se utiliza un sistema operativo Linux Ubuntu 16.10+</li><li>• La versión utilizada de PostgreSQL es 9.6</li></ul>
REQUERIMIENTOS	<ul style="list-style-type: none"><li>• NA</li></ul>
ESPECIFICACIONES	<ul style="list-style-type: none"><li>• Se requiere modificar archivos de configuración para permitir conexiones desde el servidor de aplicaciones.</li></ul>