



**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA ACADÉMICA**  
**DIRECCIÓN DE EDUCACIÓN SUPERIOR**



**PROGRAMA SINTÉTICO**

**UNIDAD ACADÉMICA:** UNIDAD PROFESIONAL INTERDISCIPLINARIA DE INGENIERÍA CAMPUS COAHUILA (UPIIC), ESCUELA SUPERIOR DE CÓMPUTO (ESCOM), UNIDAD PROFESIONAL INTERDISCIPLINARIA DE INGENIERÍA CAMPUS TLAXCALA (UPIIT)

**PROGRAMA ACADÉMICO:** Licenciatura en Ciencias de Datos

**UNIDAD DE APRENDIZAJE:** Temas selectos de procesamiento de lenguaje natural

**SEMESTRE:** VI, VII

**PROPÓSITO DE LA UNIDAD DE APRENDIZAJE**

Desarrolla aplicaciones de procesamiento de lenguaje natural a partir de sistemas de traducción, procesos de normalización, y sistemas de diálogo.

|                                   |  |            |   |  |
|-----------------------------------|--|------------|---|--|
| <b>CONTENIDOS:</b>                | I. Métodos avanzados de normalización de textos<br>II. Traducción automática<br>III. Sistemas de diálogo<br>IV. Aplicaciones de lenguaje natural |            |   |  |
| <b>ORIENTACIÓN DIDÁCTICA:</b>     | <b>Métodos de enseñanza</b>  |            | <b>Estrategias de aprendizaje</b>   |  |
|                                   | a) Inductivo   | X          | a) Estudio de casos   | X  |
|                                   | b) Deductivo   | X          | b) Aprendizaje basado en problemas  |  |
|                                   | c) Analógico   |            | c) Aprendizaje orientado proyectos  |  |
|                                   | d) Heurístico  | X          |   |  |
| <b>EVALUACIÓN Y ACREDITACIÓN:</b> | Diagnóstica  | X          | Saberes Previamente Adquiridos  | X  |
|                                   | Solución de casos  | X          | Organizadores gráficos  |  |
|                                   | Problemas resueltos  |            | Problemarios  |  |
|                                   | Reporte de proyectos   | X          | Reporte de seminarios   |  |
|                                   | Reportes de indagación   | X          | <b>Otras evidencias a evaluar:</b> presentación electrónica de resultados de indagación, reporte de uso de software, presentación e informe del proyecto, reporte de uso de software. |  |
|                                   | Reportes de prácticas  | X          |   |  |
| Evaluación escrita                |  |            |   |  |
|                                   | <b>Autor(es)</b>   | <b>Año</b> | <b>Título del documento</b>   | <b>Editorial / ISBN</b>                              |
|                                   | rafsky, D.& Martin,  | 2008       | <i>Speech and Language Processing</i>   | Pearson Prentice Hall / 9780131873216                |
| <b>BIBLIOGRAFÍA BÁSICA:</b>       | Poibeau T.   | 2017       | <i>Machine Translation</i>  | MIT Press Essential Knowledge series / 9780262534215 |
|                                   | Sarkar, D  | 2016       | <i>Text Analytics with Python</i>   | Apress / 9781484243534                               |
|                                   | Shevat, A.   | 2017       | <i>Designing Bots</i>   | O'Reilly / 9781491974827                             |
|                                   | Vajjala, S., Majumder, B., Gupta, A., & Surana, H.   | 2020       | <i>Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP System</i>   | O'Reilly Media / 9781492054054                       |





**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA ACADÉMICA**  
**DIRECCIÓN DE EDUCACIÓN SUPERIOR**



**PROGRAMA DE ESTUDIOS**

UNIDAD DE APRENDIZAJE: Temas selectos de procesamiento de lenguaje natural **HOJA 2 DE 8**

|   |  |                                   |
|---|--|-----------------------------------|
| <b>UNIDAD ACADÉMICA:</b> UNIDAD PROFESIONAL INTERDISCIPLINARIA DE INGENIERÍA CAMPUS COAHUILA (UPIIC), ESCUELA SUPERIOR DE CÓMPUTO (ESCOM), UNIDAD PROFESIONAL INTERDISCIPLINARIA DE INGENIERÍA CAMPUS TLAXCALA (UPIIT)  |  |                                   |
| <b>PROGRAMA ACADÉMICO:</b> Licenciatura en Ciencia de Datos   |  |                                   |
| <b>SEMESTRE:</b><br>VI, VII   | <b>ÁREA DE FORMACIÓN:</b><br>Profesional | <b>MODALIDAD:</b><br>Escolarizada |
| <b>TIPO DE UNIDAD DE APRENDIZAJE:</b><br>Teórica-Práctica/ Optativa   |  |                                   |
| <b>VIGENTE A PARTIR DE:</b><br>Agosto 2022/ Enero 2023  | <b>CRÉDITOS:</b>                         |                                   |
|   | <b>TEPIC:</b> 7.5                        | <b>SATCA:</b> 6.3                 |
| <b>INTENCIÓN EDUCATIVA</b>  |  |                                   |
| <p>La presente unidad contribuye al perfil de egreso del Licenciado en Ciencia de Datos con el desarrollo de habilidades en la implementación de sistemas de análisis de personalidad y opinión y de procesamiento de lenguaje natural hablado y escrito para la extracción de información, conocimiento implícito y complejo, potencialmente útil y novedoso que soporte la toma de decisiones en las organizaciones, participando en dinámicas de trabajo colaborativo e interdisciplinario con sentido ético y responsabilidad social. Asimismo, fomenta las habilidades transversales de pensamiento crítico, trabajo en equipo, comunicación efectiva, resolución de problemas, creatividad e ingenio.</p> <p>Esta unidad se relaciona de manera precedente con Estadística, Aprendizaje de máquina e Inteligencia Artificial y Matemáticas avanzadas para la ciencia de datos, de manera lateral con Análisis de series de tiempo y de manera consecuente con Big Data, Temas selectos de procesamiento de lenguaje natural</p> |  |                                   |
| <b>PROPÓSITO DE LA UNIDAD DE APRENDIZAJE</b>  |  |                                   |
| Desarrolla aplicaciones de procesamiento de lenguaje natural a partir de sistemas de traducción, procesos de normalización, y sistemas de diálogo.  |  |                                   |

|  |  |  |
|--|--|--|
| <b>TIEMPOS ASIGNADOS</b><br><br><b>HORAS TEORÍA/SEMANA:</b> 3.0<br><br><b>HORAS PRÁCTICA/SEMANA:</b> 1.5<br><br><b>HORAS TEORÍA/SEMESTRE:</b> 54.0<br><br><b>HORAS PRÁCTICA/SEMESTRE:</b> 27.0<br><br><b>HORAS APRENDIZAJE AUTÓNOMO:</b> 24.0<br><br><b>HORAS TOTALES/SEMESTRE:</b> 81.0 | <b>UNIDAD DE APRENDIZAJE DISEÑADA POR:</b><br><br>Comisión de Diseño del Programa Académico                              | <b>AUTORIZADO Y VALIDADO</b><br><br><br><br><br>Director de Educación Superior |
|  | <b>APROBADO POR:</b><br><br>Comisión de Programas Académicos del H. Consejo General Consultivo del IPN<br><br>27/06/2022 |  |



**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA ACADÉMICA**  
**DIRECCIÓN DE EDUCACIÓN SUPERIOR**



**UNIDAD DE APRENDIZAJE:** Temas selectos de procesamiento de lenguaje natural      **HOJA 3 DE 8**

| UNIDAD TEMÁTICA I<br>Métodos avanzados de normalización de texto   | CONTENIDO   | HORAS CON DOCENTE |      | HRS AA |
|--|---|-------------------|------|--------|
|  |   | T                 | P    |        |
| <b>UNIDAD DE COMPETENCIA</b><br><br>Implementa procesos de normalización a textos con base en la detección y corrección de errores, análisis morfológico, análisis sintáctico y normalización. | 1.1 Detección y corrección de errores                   | 3.0               | 1.5  | 1.5    |
|  | 1.2 Análisis morfológico                                | 4.5               | 3.0  | 1.5    |
|  | 1.2.1 Stemming y lematización                           |                   |      |        |
|  | 1.2.2 Expansión de contracciones y acrónimos            |                   |      |        |
|  | 1.2 Análisis sintáctico (parsing)                       | 4.5               | 3.0  | 1.5    |
|  | 1.4 Normalización de textos                             | 4.5               | 3.0  | 1.5    |
|  | 1.4.1 Representación vectorial                          |                   |      |        |
|  | 1.4.2 Obtención del vector a través de redes neuronales |                   |      |        |
|  | Subtotal  | 16.5              | 10.5 | 6.0    |

| UNIDAD TEMÁTICA II<br>Traducción automática  | CONTENIDO  | HORAS CON DOCENTE |     | HRS AA |
|--|--|-------------------|-----|--------|
|  |  | T                 | P   |        |
| <b>UNIDAD DE COMPETENCIA</b><br><br>Aplica los sistemas de traducción automática de acuerdo con sus características. | 2.1 Tipos de sistemas de traducción automática                       | 4.0               | 0.0 | 1.5    |
|  | 2.1.1 Basados en reglas  |                   |     |        |
|  | 2.1.2 Basados en estadística   |                   |     |        |
|  | 2.1.3 Basados en ejemplos  |                   |     |        |
|  | 2.1.4 Basados en redes neuronales                                    |                   |     |        |
|  | 2.1.5 Híbridos   |                   |     |        |
|  | 2.2. Caracterización de sistemas existentes de traducción automática | 3.5               | 0.0 | 1.5    |
|  | 2.2.1 Google Translate   |                   |     |        |
|  | 2.2.2 Microsoft Translator   |                   |     |        |
|  | 2.2.3 IBM Watson Language Translator                                 |                   |     |        |
| 2.2.4 Bing Translator  |  |                   |     |        |
| 2.2.5 Otros  |  |                   |     |        |
|  | Subtotal   | 7.5               | 0.0 | 3.0    |



**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA ACADÉMICA**  
**DIRECCIÓN DE EDUCACIÓN SUPERIOR**



UNIDAD DE APRENDIZAJE: Temas selectos de procesamiento de lenguaje natural HOJA 4 DE 8

| UNIDAD TEMÁTICA III<br>Sistemas de diálogo   | CONTENIDO  | HORAS CON DOCENTE |     | HRS AA |
|--|--|-------------------|-----|--------|
|  |  | T                 | P   |        |
| <b>UNIDAD DE COMPETENCIA</b><br><br>Implementa sistemas de diálogo con base en los tipos de sistemas diálogo, componentes y reconocimiento de voz. | 3.1 Tipos de sistemas dialogo  | 0.5               |     | 1.5    |
|  | 3.2. Componentes de sistemas diálogo   | 0.5               |     |        |
|  | 3.3 Herramientas de implementación de sistemas de dialogo                        | 0.5               |     |        |
|  | 3.4. Agente conversacional basado en texto escrito                               | 3.0               | 3.0 | 2.0    |
|  | 3.5. Agente conversacional basado en reconocimiento de voz y generación de habla | 4.5               | 4.5 | 2.0    |
|  | 3.5.1 Fundamentos de reconocimiento de voz<br>3.5.2 Generación de habla          |                   |     |        |
|  | Subtotal   | 9.0               | 7.5 | 5.5    |

| UNIDAD TEMÁTICA IV<br>Aplicaciones de lenguaje natural  | CONTENIDO  | HORAS CON DOCENTE |      | HRS AA |
|---|--|-------------------|------|--------|
|   |  | T                 | P    |        |
| <b>UNIDAD DE COMPETENCIA</b><br><br>Implementa aplicaciones de lenguaje natural a partir de la desambiguación, generación de respuestas, resolución de anáforas y correferencias, reconocimiento de inferencias y procesamiento de texto biomédicos | 4.1 Desambiguación de sentido de palabras              | 3.0               | 1.5  | 2.0    |
|   | 4.2 Generación de respuestas a preguntas               | 4.5               | 1.5  | 2.0    |
|   | 4.3 Resolución de anáfora y correferencia              | 4.5               | 1.5  | 2.0    |
|   | 4.4 Reconocimiento de inferencias (textual entailment) | 4.5               | 1.5  | 2.0    |
|   | 4.5 Procesamiento de textos biomédicos                 | 4.5               | 3.0  | 1.0    |
|   |  | Subtotal          | 21.0 | 9.0    |



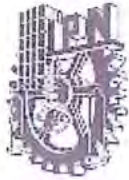
**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA ACADÉMICA**  
**DIRECCIÓN DE EDUCACIÓN SUPERIOR**



UNIDAD DE APRENDIZAJE: Temas selectos de procesamiento de lenguaje natural      HOJA 5 DE 8

| ESTRATEGIAS DE APRENDIZAJE  | EVALUACIÓN DE LOS APRENDIZAJES   |
|---|--|
| <p><b>Estrategia de aprendizaje de Estudio de Casos</b></p> <p>El alumno desarrollará las siguientes actividades:</p> <ol style="list-style-type: none"><li>1. Desarrollo de conceptos teóricos e indagación documental.</li><li>2. Presentación oral de los resultados de indagación documental.</li><li>3. Realización de prácticas en equipo e individual de los temas que sean vistos en clase que le permita el análisis de los temas.</li><li>4. Resolución de casos de estudio en equipo e individual de los temas que sean vistos en clase que le permita el análisis de los temas.</li><li>5. Desarrollo de un proyecto donde se implemente los procesos de lenguaje natural vistos en clase.</li><li>6. Presentación oral del proyecto</li><li>7. Uso de software para procesamiento de textos en lenguaje natural.</li></ol> | <p>Evaluación diagnóstica.</p> <p>Portafolio de evidencias:</p> <ol style="list-style-type: none"><li>1. Reporte de indagación</li><li>2. Presentación electrónica de resultados de indagación</li><li>3. Reporte de prácticas</li><li>4. Reporte de solución de casos de estudio</li><li>5. Reporte de proyecto</li><li>6. Presentación e informe de proyecto</li><li>7. Reporte de uso de software</li></ol> |

| RELACIÓN DE PRÁCTICAS |   |                            |                      |
|-----------------------|---|----------------------------|----------------------|
| PRÁCTICA No.          | NOMBRE DE LA PRÁCTICA   | UNIDADES TEMÁTICAS         | LUGAR DE REALIZACIÓN |
| 1                     | Normalización de textos   | I                          | Sala de cómputo      |
| 2                     | Caracterización de un sistema existente de traducción automática  | II                         |                      |
| 3                     | Implementación de un agente conversacional                        | III                        |                      |
| 4                     | Implementación de desambiguación de sentido de palabras           | IV                         |                      |
| 5                     | Implementación de resolución de anáfora                           | IV                         |                      |
| 6                     | Caracterización de un sistema existente de respuestas a preguntas | IV                         |                      |
| 7                     | Implementación de reconocimiento de inferencias                   | IV                         |                      |
| 8                     | Procesamiento de textos biomédicos                                | IV                         |                      |
|                       |   | <b>TOTAL DE HORAS 27.0</b> |                      |



**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA ACADÉMICA**  
**DIRECCIÓN DE EDUCACIÓN SUPERIOR**



UNIDAD DE APRENDIZAJE: Temas selectos de procesamiento de lenguaje natural HOJA 6 DE 8

| Bibliografía |  |      |  |  |           |            |       |
|--------------|--|------|--|--|-----------|------------|-------|
| Tipo         | Autor(es)  | Año  | Título del documento   | Editorial  | Documento |            |       |
|              |  |      |  |  | Libro     | Antologías | Otros |
| C            | Bengfort, B., Bilbro, R., & Ojeda, T.              | 2018 | <i>Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning</i>  | O'Reilly Media, Inc./ 9781491963043                  | X         |            |       |
| C            | Bird, S., Klein, E., Loper, E.                     | 2009 | <i>Natural Language Processing with Python</i>   | O'Reilly/ 9780596516499                              | X         |            |       |
| C            | Ganegedara, T.                                     | 2018 | <i>Natural Language Processing with TensorFlow: Teach Language to Machines Using Python's Deep Learning Library</i>                                | Packt Publishing Ltd./ 9781788478311                 | X         |            |       |
| C            | Ghosh, S., & Gunning, D.                           | 2019 | <i>Natural Language Processing Fundamentals: Build Intelligent Applications that can Interpret the Human Language to Deliver Impactful Results</i> | Packt Publishing Ltd./ 9781789954043                 | X         |            |       |
| C            | Gomez-Perez, J. M., Denaux, R., & Garcia-Silva, A. | 2020 | <i>A Practical Guide to Hybrid Natural Language Processing: Combining Neural Models and Knowledge Graphs for NLP</i>                               | Springer Nature/ 9783030448295                       | X         |            |       |
| B            | Jurafsky, D., Martin, J.                           | 2008 | <i>Speech and Language Processing</i>  | Pearson Prentice Hall / 9780131873216                | X         |            |       |
| B            | Poibeau T.   | 2017 | <i>Machine Translation</i>   | MIT Press Essential Knowledge series / 9780262534215 | X         |            |       |
| B            | Sarkar, D.   | 2019 | <i>Text Analytics with Python:</i>   | Apress/ 9781484243534                                | X         |            |       |
| B            | Shevat, A.   | 2017 | <i>Designing Bots</i>  | O'Reilly / 9781491974827                             | X         |            |       |
| B            | Vajjala, S., Majumder, B., Gupta, A., Surana, H.   | 2020 | <i>Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems</i>   | O'Reilly Media/ 9781492054054                        | X         |            |       |

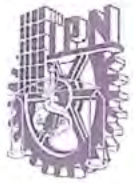


**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA ACADÉMICA**  
**DIRECCIÓN DE EDUCACIÓN SUPERIOR**



UNIDAD DE APRENDIZAJE: Temas selectos de procesamiento de lenguaje natural HOJA 7 DE 8

| Recursos digitales  |                       |   |                            |                                      |                       |  |   |                  |
|---|-----------------------|---|----------------------------|--------------------------------------|-----------------------|--|---|------------------|
| Autor, año, título y Dirección Electrónica  | T<br>e<br>x<br>t<br>o | S<br>i<br>m<br>u<br>l<br>a<br>d<br>o<br>r | I<br>m<br>a<br>g<br>e<br>n | T<br>u<br>t<br>o<br>r<br>i<br>a<br>l | V<br>i<br>d<br>e<br>o | P<br>r<br>e<br>s<br>e<br>n<br>t<br>a<br>c<br>i<br>o<br>n | D<br>i<br>c<br>c<br>i<br>o<br>n<br>a<br>r<br>i<br>o | O<br>t<br>r<br>o |
| Procesamiento de lenguaje natural. Recuperado el 29 de septiembre de 2020 <a href="https://www.sas.com/es_ar/insights/analytics/what-is-natural-language-processing-nlp.html">https://www.sas.com/es_ar/insights/analytics/what-is-natural-language-processing-nlp.html</a> .   | X                     |   |                            |                                      |                       |  |   |                  |
| El Procesamiento de Lenguaje Natural (PLN) en la producción de noticias. Recuperado el 29 de septiembre de 2020 <a href="https://www.youtube.com/watch?v=pPmpHyT_fCA..">https://www.youtube.com/watch?v=pPmpHyT_fCA..</a>   |                       |   |                            |                                      | X                     |  |   |                  |
| Aplicaciones de procesamiento de lenguaje natural. Recuperado el 29 de septiembre de 2020 <a href="https://core.ac.uk/download/pdf/18586869.pdf">https://core.ac.uk/download/pdf/18586869.pdf</a>   | X                     |   |                            |                                      |                       |  |   |                  |
| ¿Qué es la traducción automática? Recuperado el 29 de septiembre de 2020 <a href="https://www.systransoft.com/es/systran/tecnologia/que-es-la-traduccion-automatica/">https://www.systransoft.com/es/systran/tecnologia/que-es-la-traduccion-automatica/</a> .<br>Visto   | X                     |   |                            |                                      |                       |  |   |                  |
| Sistemas de diálogo hablado y multimodal. Recuperado el 24 de mayo de 2021 <a href="http://www.ugr.es/~rlopezc/sistemas_dialogo.htm">http://www.ugr.es/~rlopezc/sistemas_dialogo.htm</a> .  | X                     |   |                            |                                      |                       |  |   |                  |
| Python implementations of word sense disambiguation. Recuperado el 24 de mayo de 2021 <a href="https://github.com/alvations/pywsd">https://github.com/alvations/pywsd</a>   |                       |   |                            |                                      |                       |  |   | X                |
| How to build an open-domain question answering system?. Recuperado el 24 de mayo de 2021 <a href="https://lilianweng.github.io/lil-log/2020/10/29/open-domain-question-answering.html">https://lilianweng.github.io/lil-log/2020/10/29/open-domain-question-answering.html</a>  | X                     |   |                            |                                      |                       |  |   |                  |
| ¿Quién habla? Agente conversacional vs chatbot vs asistente virtual. Recuperado el 29 de septiembre de 2020 <a href="https://searchdatacenter.techtarget.com/es/cronica/Quien-habla-Agente-conversacional-vs-chatbot-vs-asistente-virtual">https://searchdatacenter.techtarget.com/es/cronica/Quien-habla-Agente-conversacional-vs-chatbot-vs-asistente-virtual</a> . | X                     |   |                            |                                      |                       |  |   |                  |
| Anaphora resolution Recuperado el 24 de mayo de 2021 <a href="https://es.slideshare.net/kancho/anaphora-resolution">https://es.slideshare.net/kancho/anaphora-resolution</a>  | X                     |   |                            |                                      |                       |  |   |                  |
| Textual entailment with TensorFlow Recuperado el 24 de mayo de 2021 <a href="https://www.oreilly.com/content/textual-entailment-with-tensorflow/">https://www.oreilly.com/content/textual-entailment-with-tensorflow/</a>   | X                     |   |                            |                                      |                       |  |   |                  |
| Biomedical text natural language processing (BioNLP) using scispaCy Recuperado el 24 de mayo de 2021 <a href="https://qbnegrini.com/post/biomedical-text-nlp-scispacy-named-entity-recognition-medical-records/">https://qbnegrini.com/post/biomedical-text-nlp-scispacy-named-entity-recognition-medical-records/</a>  | X                     |   |                            |                                      |                       |  |   |                  |



**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA ACADÉMICA**  
**DIRECCIÓN DE EDUCACIÓN SUPERIOR**



**UNIDAD DE APRENDIZAJE:** Temas selectos de procesamiento de lenguaje natural    **HOJA 8 DE 8**

**PERFIL DOCENTE:** Ingeniería, Licenciatura y/o Maestría en Ciencias Físico Matemáticas o afines

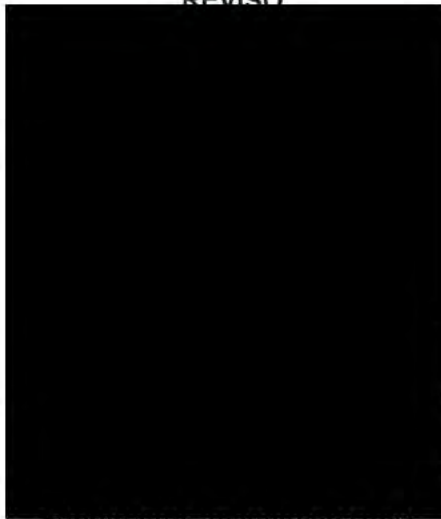
| <b>EXPERIENCIA PROFESIONAL</b>  | <b>CONOCIMIENTOS</b>   | <b>HABILIDADES DIDÁCTICAS</b>  | <b>ACTITUDES</b>  |
|---|--|--|---|
| Preferentemente dos años en la profesión en el área de Inteligencia Artificial y/o Procesamiento de lenguaje natural y dos años de docencia a nivel superior. | En Inteligencia Artificial y Procesamiento de lenguaje natural.<br>Del Modelo Educativo Institucional. | Discursivas<br>Cognoscitivas<br>Metodológicas<br>De conducción del grupo<br>Para evaluar | Responsabilidad<br>Tolerancia<br>Honestidad<br>Respeto<br>Paciencia<br>Disciplina<br>Constancia |

**ELABORÓ**



Dra. Olga Kolesnikova  
**Coordinadora**

**REVISÓ**



**Subdirector Académico UPIN**



Dr. Eric Manuel Rosales Peña Alfaro  
**Participante**

