



COLECCIÓN DE TESIS Y PROPUESTA DE INVESTIGACIÓN EN TICs: UN RECURSO PARA SU ANÁLISIS Y ESTUDIO

SAMUEL GONZÁLEZ-LÓPEZ

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA
sgonzalez@inaoep.mx

AURELIO LÓPEZ-LÓPEZ

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA
alopez@inaoep.mx

RESUMEN

La culminación de una carrera en Instituciones de Educación Superior, en algunos casos, implica la redacción de una tesis. Durante este proceso el estudiante expresa sus ideas en un documento, que es posteriormente revisado por un asesor académico. Este documento posee características establecidas, sin embargo, en ocasiones presentan deficiencias que demandan mayor tiempo de revisión. Nuestro trabajo aporta una Colección de Tesis y Propuestas de Investigación (CoITyPI 1.0) como un recurso tecnológico para complementar la formación del estudiante en asignaturas como metodología de la investigación. Además, la colección puede ser explorada por investigadores que buscan identificar comportamientos propios de los elementos de una tesis. También se describen las principales tareas desarrolladas a partir de la colección, así como dos sub-colecciones anotadas manualmente.

Palabras clave: Análisis de Textos, Instituciones de Educación Superior, TICs, Redacción de Textos, Tesis de Grado.





INTRODUCCIÓN

En México la investigación y desarrollo de nuevo conocimiento es soportado en su mayoría por los Centros de Investigación y Universidades, con esquemas ya sea público o privado. El crecimiento del ingreso de alumnos a niveles posgrado ha sido ascendente, del 2010 al 2013 el incremento fue del 36%, persistiendo un problema fuerte que afecta a todas las instituciones, el porcentaje de alumnos titulados. Este indicador a nivel posgrado se ubica alrededor del 64% según reportes de ANUIES¹. A nivel Licenciatura este indicador fue del 71% para el periodo 2012-2013. En este nivel los alumnos tienen diferentes modalidades de egreso, entre ellas la elaboración de una tesis. Para el nivel Técnico Superior Universitario (TSU), el porcentaje de titulados en el mismo periodo que Licenciatura fue del 60%, en donde los alumnos culminan su programa con la elaboración de una tesina.

Los factores para lograr la eficiencia terminal son diversos. En el estudio de (Muñoz, 2011) el 8% se debió a problemas con la estructuración del documento, aunado a la falta de asesoría y dificultad en la definición del tema a desarrollar, entre otros. Un estudio sobre la dificultad para escribir la sección conclusión de una tesis, reveló que los estudiantes no tenían certeza del contenido y organización de una conclusión (Bitchener & Basturkmen, 2006).

Motivados por generar recursos tecnológicos que puedan ayudar a los estudiantes a mejorar su escritura, en este trabajo presentamos una Colección de Tesis y Propuestas de Investigación (CoITyPI 1.0) en el área de computación y tecnologías de la información. El objetivo principal es ayudar a los estudiantes en las etapas tempranas de la redacción del borrador de tesis. Esta colección provee varias alternativas de descarga y uso, detalladas en capítulos posteriores. El trabajo que aquí se presenta es parte de un proyecto titulado “Análisis lingüístico de borradores de propuestas de investigación de estudiantes de Pregrado”, el cual tiene como objetivo generar métodos que permitan analizar de forma automática la estructura de una tesis.

¹Asociación Nacional de Universidades e Instituciones de Educación Superior





ANTECEDENTES

Actualmente el acceso a documentos institucionales se ha incrementado gracias a repositorios que brindan un acceso libre, o en ocasiones restringido, a los propios estudiantes de esa institución. La divulgación de documentos tipo tesis, es un elemento importante para el crecimiento académico de nuevas investigaciones y de la diversidad de soluciones que pudieran considerarse para afrontar un problema real. Cabe mencionar, que en el nivel posgrado muchas tesis se publican en conferencias o revistas nacionales e internacionales. Sin embargo, muchas de las tesis de nivel pregrado (Licenciatura y TSU), no logran tener una difusión más allá de sus instituciones.

Durante la revisión de colecciones de tesis disponibles en internet encontramos las siguientes:

La Biblioteca Central de la UNAM², que provee una amplia colección de tesis.

El Instituto Politécnico Nacional³ proporciona un repositorio para la descarga de tesis.

La Colección de tesis digitales, de la Universidad de las Américas Puebla⁴.

Además existe la Red Mexicana de Repositorios Institucionales (REMERI)⁵, donde se encuentra un listado para el acceso a repositorios de instituciones públicas y privadas. Estos repositorios brindan acceso a tesis de las carreras que imparten las instituciones. Una de las diferencias entre las colecciones mencionadas y nuestra colección, es la manera en la que se puede acceder a los elementos de una tesis. Por ejemplo, se puede realizar la descarga del elemento objetivo general de los niveles de posgrado, de pregrado o ambos. Otra diferencia es que la colección es especializada en computación y tecnologías de la información. Además se cuenta con dos sub-colecciones anotadas por instructores con experiencia en revisión de tesis.

Para el español y específicamente en documentos tipo tesis no se han encontrado recursos que provean textos con algún tipo de pre-procesamiento para su análisis. Sin embargo, se han detectado otros tipos de documentos como ensayos de estudiantes de

² <http://bc.unam.mx/index-alterno.html>

³ http://www.dirbibliotecas.ipn.mx/Paginas/Tesis_Electronicas.aspx

⁴ http://catarina.udlap.mx/u_dl_a/tales/

⁵ <http://www.remeri.org.mx/repositorios/>





segunda lengua, principalmente en inglés, como el que provee la Universidad de Uppsala, Suecia⁶.

LA COLECCIÓN

Los documentos recolectados que integran la colección, corresponden a tesis y propuestas de investigación escritas por estudiantes y asciende a 468 documentos. Cada elemento de la colección ha sido revisado en algún momento por un comité de tesis. Dos tipos de documentos se distinguen: el nivel posgrado (Doctorado y Maestría) y nivel pregrado (Licenciatura y TSU).

Los documentos fueron recolectados de repositorios de universidades públicas y centros de investigación. El 97% corresponden a instituciones mexicanas y el 3% a universidades de otros países hispanohablantes. Cabe mencionar que se descargaron los documentos completos en formato Word o Pdf. En algunos casos, los proyectos de investigación desarrollados por los estudiantes fueron considerados como parte de la colección. Algunas tesis estaban segmentadas en varios archivos, por lo que se agruparon en una sola carpeta.

Los elementos extraídos fueron: Título, Planteamiento del Problema, Justificación, Preguntas de Investigación, Hipótesis, Objetivos, Metodología y Conclusiones. La selección de los elementos se realizó considerando las recomendaciones de los autores de libros de metodología (Hernández, et al. 2010).

La recolección y la extracción fueron realizadas de manera manual e identificando solo aquellas secciones que estaban delimitadas por un título. La siguiente tabla muestra la cantidad de elementos por nivel.

| Grado | Elementos |
|--------------|-----------|
| Doctorado | 59 |
| Maestría | 181 |
| Licenciatura | 150 |
| TSU | 78 |
| Total | 468 |

⁶

http://www.engelska.uu.se/Forskning/engelsk_sprakvetenskap/Forskningsomraden/Electronic_Resource_Projects/





Tabla 1. Colección ColTyPI 1.0

Se puede observar que los grados de Doctorado y TSU poseen menor cantidad de elementos, comparados con los grados de Maestría y Licenciatura. Para el grado de doctorado, la producción es menor debido a su propia naturaleza, mientras que en el grado TSU los repositorios disponibles son pocos.

Sin embargo, la cantidad de elementos para el nivel posgrado es de 240 y 228 para el nivel pregrado, evidenciando un equilibrio entre ambos niveles (ver Tabla 2).

| Colección | | | |
|----------------------------|----------|----------|-------|
| Sección | Posgrado | Pregrado | Total |
| Planteamiento del problema | 132 | 100 | 232 |
| Justificación | 108 | 132 | 240 |
| Preguntas de investigación | 133 | 19 | 152 |
| Hipótesis | 71 | 26 | 97 |
| Objetivos | 218 | 212 | 430 |
| Metodología | 113 | 71 | 184 |
| Conclusiones | 216 | 197 | 413 |
| Título | 240 | 228 | 468 |
| Total | 1231 | 985 | 2216 |

Tabla 2. Colección detallada.

La cantidad total que muestra la Tabla 2 es de 2.216 elementos extraídos de la colección. El nivel de posgrado tiene 1231 elementos y el nivel pregrado tiene 895. Algunos elementos no fueron identificados en el documento, únicamente el título fue el elemento que estuvo presente en todos los textos, seguido del objetivo general.

La colección ColTyPI 1.0 se encuentra disponible en el sitio www.coltypi.org y habilitada para su descarga. Se proveen dos opciones de descarga: la primera permite al usuario descargar los archivos completos de cada tesis, la segunda permite realizar descargas personalizadas dependiendo del elemento que se desee obtener.





Esta colección ha sido utilizada como un recurso en el análisis del comportamiento de algunos elementos como objetivos y conclusiones. Además de las dos opciones de descarga, se proveen dos sub-colecciones anotadas con propósitos específicos.

Para realizar la tarea de anotado se integró un grupo de 4 anotadores con experiencia en la revisión de tesis del área de computación (profesores de universidades) para crear un conjunto de elementos de referencia.

La primera sub-colección se enfocó en anotar 15 conclusiones de Licenciatura y 15 elementos de TSU. Los elementos marcados por los anotadores, considerando el objetivo y la conclusión de una tesis fueron: conexión del objetivo con las conclusiones (cobertura), evidencia de juicios y opiniones (opinión), y evidencia de trabajo futuro (especulación). Estas características son señaladas por autores de libros en metodología de la investigación.

La segunda sub-colección fue marcada considerando las preguntas metodológicas: ¿Qué se va hacer?, ¿Para qué se va hacer? y ¿Cómo se va hacer?. Fueron etiquetados 300 objetivos por cada anotador. La respuesta a estas tres preguntas metodológicas, forman parte del grupo de referencia para la evaluación de objetivos.

Las dos sub-colecciones pueden ser usadas para ejemplificar a los estudiantes diferentes tipos de casos, en asignaturas de metodología de la investigación. También por investigadores que deseen explorar, bajo técnicas de inteligencia artificial (por ejemplo, técnicas de procesamiento de lenguaje natural o aprendizaje) la identificación de alguno de los elementos. Cada uno de las sub-colecciones y las instrucciones de anotado pueden ser descargados del sitio www.coltypi.org. También se presentan los niveles de acuerdo entre los anotadores en cada sub-colección. Además, se encuentra disponible una herramienta de marcado online utilizada en la segunda sub-colección.

ANÁLISIS RELACIONADOS A COLTYPI 1.0

La colección ha sido utilizada en el análisis lingüístico de propuestas de investigación de estudiantes de grado Licenciatura y TSU. Las características que se han investigado son:

Riqueza léxica: la habilidad del estudiante para entender diferentes conceptos, que les permiten estructurar oraciones ricas en vocabulario y significado. En esta tarea se analizaron los dos niveles: pregrado y posgrado, además se realizó una prueba piloto (González & López-López, 2015).

Dificultad en la redacción de las conclusiones: los elementos analizados involucran la conclusión y el objetivo general de la propuesta (González-López, & López-López, 2015).





Coherencia: Definida como la conexión de todas las partes de un texto en un todo, se examinó coherencia en tesis y propuestas de investigación (López & López-López, 2014).

Sentencias débiles: la sección conclusiones fue analizada para identificar frases que no seguían pautas metodológicas previamente establecidas (González et al. 2014).

Flujo de conceptos: Se propusieron métodos para identificar las transiciones entre los párrafos que integran una conclusión (González-López & López-López, 2014).

CÓMO ACCEDER A LA COLECCIÓN

El sitio web donde se aloja la colección integra un menú con seis opciones. La opción Corpus despliega el submenú Descargas. En esta sección se puede descargar en un archivo de formato “rar” las colecciones completas, es decir, los archivos que contienen las tesis y propuestas de investigación. Además están disponibles cuatro archivos: Doctorado.rar, Maestria.rar, Licenciatura.rar y TSU.rar.

En el menú Consultas se despliega un recuadro con el grado y las secciones que están disponibles para su descarga, según se muestra en la siguiente imagen:





CoTyPI 1.0
Colección de Tesis y Propuestas de Investigación

Inglés

Inicio Corpus **Consultas** Corpus etiquetado Contactanos Herramientas

INADE
Instituto Nacional de Astrofísica,
Óptica y Electrónica

Elija los criterios de su búsqueda

Grado **Secciones**

Todos los Grados Limpiar selección Grados Todos las secciones Limpiar selección de secciones

Doctorado Titulo Metodología

Maestría Problema Resultado

Licenciatura Objetivo Tipo

TSU Preguntas Estatus

Hipótesis Otros

Justificación

Generar XML

Figura 1. Pantalla de la opción Consultas

El usuario tiene la posibilidad de marcar aquellas casillas que sean de su interés. El resultado es un archivo en formato XML, dando portabilidad y fácil acceso a la colección. La descarga puede ser seleccionando todas las secciones (por default aparecen marcadas) o elegir alguna de ellas. En el grado a seleccionar se sugiere marcar alguno de ellos. Cabe remarcar que algunas tesis no tienen todos los elementos. En la siguiente tabla se muestra un ejemplo del archivo XML (omitiendo etiquetas propias de XML):





```
<Titulo>Metodología de Medición y Evaluación de la Usabilidad en Sitios Web
Educativos</Titulo>
  <Problema>La gran cantidad de información disponible en Internet, la complejidad y
  heterogeneidad de.....</Problema>
  <Metodologia>Es una metodología que soporta la evaluación remota de sitios
  Web [PAT99]....</Metodologia>
```

Tabla 3. Ejemplo de archivo XML

El archivo puede ser utilizado para propósitos de experimentación, así la lectura del archivo puede ser programada de manera automática. Sin embargo, para usuarios que solo deseen visualizar en pantalla el archivo, se puede recurrir a lectores de archivos de formato XML⁷. El formato de codificación de los caracteres especiales es utf-8. En el menú Corpus etiquetado, se despliega un submenú con dos opciones Conclusiones-objetivos y Objetivos.

Conclusiones-objetivos:

La primera opción nos permite descargar la colección anotada en archivo Word, de cada uno de los anotadores, además de las instrucciones para realizar la anotación. Cada una de las conclusiones están marcadas por el anotador. En la Tabla 4, se puede observar en color amarillo el texto que se conecta con el objetivo general, mientras que en color verde fue marcado el trabajo futuro. En la parte derecha de la tabla se muestra el grado de opinión identificado por el anotador. Las primeras 15 conclusiones corresponden al grado TSU y las siguientes corresponden al grado Licenciatura.

⁷ <http://codebeautify.org/xmlviewer>





| | |
|---|--|
| 1. Objetivo: Desarrollar un sistema de control y monitoreo de encendido de luz en áreas comunes a través de un controlador de lógica programable (PLC). | |
| Conclusión: El conocimiento y manejo del software del PLC y de LabVIEW se realizó de una manera satisfactoria, ya que el conocer estas herramientas de software permitió diseñar y construir una interfaz de usuario bastante eficiente y amigable, así como diseñar un programa para control del PLC. Se logró realizar la comunicación entre el software (LabVIEW) y el hardware (PLC), para minimizar el uso de energía de los laboratorios, cubículos y áreas comunes presentadas. Por lo que el sistema de Control de energía en base a PLC presentado cumple los objetivos propuestos, así como también minimiza el uso de energía, es amigable al usuario y puede ser expandido a múltiples cubículos, laboratorios y áreas comunes. | Escala para Opinión Si, Mucho Si, Poco No hay Opinión |

Tabla 4. Conclusión-objetivo etiquetada

Como se mencionó esta sub-colección presenta 30 casos etiquetados, los cuales pueden usarse como un recurso complementario a los libros de metodología. Esta colección fue etiquetada primero por dos anotadores, obteniendo el nivel de acuerdo Cohen Kappa (Landis & Koch, 1977) de: 0.92 (Casi perfecto) para la característica Cobertura, 0.65 (Considerable) para el elemento Especificación y para la Opinión: Si, mucho 0.37 (Aceptable) Si, Poco 0.07 (Leve) y No hay Opinión 0.01 (Leve). Posteriormente otro par del grupo de anotadores etiquetó la característica Opinión, obteniendo: Si, mucho 0.47 (Moderado), 0.21 (Aceptable), and 0.44 (Moderado). Este corpus fue estudiado en (González-López, & López-López, 2015).

Objetivos: La segunda opción provee un corpus de 300 objetivos marcados por 4 anotadores. Para esta tarea se utilizó la herramienta de anotación, la cual permite al anotador realizar el marcado en línea. A continuación se muestra un ejemplo de un objetivo anotado:





Objetivo:

Generar una herramienta de soporte para el estudio del álgebra en el nivel medio superior a través de un sistema computacional que administre los objetos de aprendizaje con IEEE-LOM asociados con diversos tópicos de esta disciplina.

¿Qué se va hacer?: *Generar una herramienta de soporte*

¿Para qué se va hacer?: *para el estudio del álgebra en el nivel medio superior*

¿Cómo se va hacer?: *a través de un sistema computacional que administre los objetos de aprendizaje con IEEE-LOM asociados con diversos tópicos de esta disciplina.*

Tabla 5. Ejemplo de objetivo etiquetado

La sub-colección de 300 objetivos corresponde a objetivos de los niveles de pregrado y posgrado. Sin embargo, se tiene la información necesaria para identificar cada grado. Los archivos descargables que se presentan como disponibles son los siguientes:

Guía para etiquetado- Formato Word.

Anotador 1- Formato Excel.

Anotador 2- Formato Excel.

Anotador 3- Formato Excel.

Anotador 4- Formato Excel.

300 objetivos etiquetados- Formato Excel.

Cada archivo de los anotadores contiene 3 columnas, mientras que el archivo de los 300 objetivos muestra una columna con el objetivo sin etiquetar. Los niveles de acuerdo Fleiss Kappa obtenidos entre los 4 anotadores son los siguientes: ¿Qué se va hacer? 0.629 (Considerable), ¿Para qué se va hacer? 0.586 (Moderado), ¿Cómo se va hacer? 0.601 (Considerable).

Herramienta de Etiquetado: En el menú Herramientas se muestra el submenú Etiquetador de Objetivos el cual envía a una pantalla de acceso con un usuario demo para explorar la herramienta. A continuación se muestra la herramienta de etiquetado:





CoTyPI 1.0
Colección de Tesis y Propuestas de Investigación

English Salir

HERRAMIENTA DE ETIQUETADO



Objetivos de tesis

| Archivos existentes | | | |
|------------------------------|----------|----------|--|
| N/E(Respuesta no encontrada) | | | |
| # | Id_tesis | Grado | Objetivo |
| 1 | 1 | posgrado | El objetivo general de este proyecto de investigación es: Adaptar la metodología de desarrollo arquitectónico del SEI para que pueda ser empleada por equipos de desarrollo pequeños |

Preguntas Metodológicas

¿Qué se va a hacer?

¿Para qué se va a hacer?

¿Cómo se va a hacer?

Borrar Todo

Guardar datos

Datos Guardados

| id | ¿Qué se va a hacer? | ¿Para que se va a hacer? | ¿Cómo se va a hacer? |
|----|---------------------|--------------------------|----------------------|
|----|---------------------|--------------------------|----------------------|

Figura 2. Herramienta de etiquetado

Actualmente la herramienta está diseñada para anotar objetivos, pero puede ser personalizada dependiendo de las necesidades. De esta manera la herramienta puede ser utilizada previa solicitud con los desarrolladores. El usuario y contraseña para el acceso demo se realiza bajo el término “coltypidemo”.

AGRADECIMIENTOS

Agradecemos a los anotadores N. Angélica Álvarez Torres, J. Miguel García Gorrostieta, Indelfonso Rodríguez Espinoza, y Mariela Altamirano. A los estudiantes que realizaron sus prácticas profesionales participando en la creación de la Colección. Además, a Abril Gloria R. desarrolladora del sitio web de la colección y de la herramienta de etiquetado. Esta colección forma parte de los productos obtenidos de la tesis doctoral del primer autor, el cual recibió





apoyo por parte de CONACYT con la beca 1124002. El segundo autor fue apoyado parcialmente por el SNI, México.

CONCLUSIONES

Comunicar ideas a través del lenguaje escrito es esencial para la sociedad del conocimiento, como también para los estudiantes cuando redactan su tesis. Algunas estrategias se han propuesto e implementado para mejorar la escritura de los estudiantes. Estas van desde aquellas diseñadas para mejorar la didáctica, hasta las que integran el uso de la tecnología, como tutores inteligentes y plataformas como Moodle para la gestión de los documentos.

La colección que se presenta en este trabajo es un recurso tecnológico que busca impactar en los estudiantes que se encuentran en el desarrollo de su propuesta de investigación o tesis, ya sea para culminar una carrera o para realizar un proyecto de investigación en alguna asignatura.

En trabajos futuros, se pretende ampliar la colección e incorporar nuevas sub-colecciones. Además, buscamos que los métodos generados en el proyecto de Análisis Lingüístico, sirvan a los estudiantes, así como la colección que se ha descrito.





BIBLIOGRAFÍA Y REFERENCIAS

- Muñoz C. 2011. Como elaborar y asesorar una investigación de tesis. Pearson ISBN: 978-607-32-0456-9.
- Bitchener, J. and Basturkmen, H. 2006. Perceptions of the difficulties of postgraduate L2 thesis students writing the discussion section. *Journal of English for Academic Purposes* 5, 4-18.
- Hernández, R., Fernández, C., y Batista, M. 2010. Metodología de la Investigación. México DF, Mc Graw Hill ISBN: 978-607-15-0291-9.
- González, S., and López-López, A. 2015. Lexical Analysis of Student Research Drafts in Computing, *Computer Applications in Engineering Education*.
- González-López, S. and López-López, A. 2015. Mining of Conclusions of Student Texts for Automatic Assessment, *The 28th International FLAIRS 2015*.
- López, SG., and López-López, A. 2014. Mining Domain Knowledge for Coherence Assessment of Students Proposal Drafts, *Educational Data Mining*. 524 pp. 229-255, Springer.
- González, S., Bethard, S., and López-López, A. 2014. Identifying Weak Sentences in Student Drafts: A Tutoring System. *Methodologies and Intelligent Systems for Technology Enhanced Learning*, 292 pp. 77-85, Springer.
- González-López, S., and López-López, A. 2014. Analysis of Concept Sequencing in Student Drafts, *9th European Conference on Technology Enhanced Learning*, pp. 422-427, Springer.
- Landis, J., Koch, G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33(1), 159–17.



