



Research Trends at UNAP: a Historical Analysis

Jose Luis Juarez Ruelas, Drake Leonel Chambilla Choquecota,
Lizbeth Mayummy Mamani Mamani and Adderly Mendoza Nina

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

July 20, 2021

Tendencias de la investigación en la UNAP: un análisis histórico

Jose Juarez
Ing. De Sistemas
Universida Nacional del Altiplano
Puno, Perú
jjuaraz@unap.edu.pe

Drake Chambilla
Ing. De Sistemas
Universida Nacional del Altiplano
Puno, Perú
drchambillac@est.unap.edu.pe

Lizbeth Mamani
Ing. De Sistemas
Universida Nacional del Altiplano
Puno, Perú
lizmamanima@est.unap.edu.pe

Adderly Mendoza
Ing. De Sistemas
Universida Nacional del Altiplano
Puno, Perú
amendozan@est.unap.edu.pe

Abstract—In the context of Big Data, where the exponential growth of data continues, mainly unstructured data, such as text documents. In this scenario, when an approach is needed that allows the analysis of text documents, the use of analysis methods grows, which allow to present in a figure, the most relevant terms that make up the documents, based on the frequency of occurrence of terms. This paper makes an analysis of the trend of research works at the National University of the Altiplano Puno, using for this the research works found in the institutional repository. The approach used here allowed to identify the positive, negative, neutral and atypical trends in relation to the research works, this from the knowledge base that has a positive evolutionary trend. (Abstract)

Keywords—data extraction, big data, research work, trend analysis (key words)

I. INTRODUCCIÓN

Recientemente, técnicas y tecnologías son usadas para extracción de información de sitios web como el repositorio de la UNAP[1]. Usualmente estos programas están relacionados con la indexación web, la que permite recopilar información de la web. Al mismo tiempo procesos como el almacenar datos están sufriendo innumerables innovaciones [2]. En ese contexto de constantes innovaciones, desafíos, emergió el concepto de Big Data, que es una de las tendencias más importantes del sector tecnológico, con el potencial para generar conocimiento a partir de grandes volúmenes de datos (Big Data)[3]. Dentro de las características comunes al concepto de Big Data, además de la velocidad de generación y del volumen, está la variedad de los datos generados, generalmente marcada por la presencia de datos no estructurados, tales como documentos de texto puro[4][7].

A fin de evaluar las tendencias de investigación en la UNAP, fue delimitado al contexto de análisis de tendencias de la investigación en la UNAP. Puesto que el repositorio institucional de la UNAP, al igual que otras universidades almacenan los trabajos de investigación (tesis) producidos a nivel pregrado y posgrado por la Universidad.

En este trabajo fueron analizados los trabajos presentes en el repositorio institucional de la UNAP, que datan aproximadamente desde el 2010 en adelante.

La literatura relata diversos abordajes para evaluar las tendencias en investigación. Hunt et al. [6] realizaron un análisis visual de tendencias a partir de los títulos de artículos publicados en cuatro revistas en el área de turismo que tienen el ranking más alto. La investigación fue realizada considerando 4 décadas de estudio histórico considerando la muestra para los años 1982, 1992, 2002 y 2012. La técnica de visualización fue la nube de palabras. Haugerud [5] realizó un análisis a los títulos y palabras clave de la revista *American Ethnologist*, analizando los datos a través de nube de palabras, en la cual fue posible identificar las palabras dominantes en una de las cuatro décadas analizadas. Ahearn [1] evidencian el uso de las palabras para poder identificar patrones de interés, y como estos temas de interés se desarrollan variando a lo largo del tiempo. Estos autores argumentan sobre la importancia del análisis de tendencias y las técnicas que permiten encontrar e identificar patrones de interés.

Con base en el análisis de la literatura relacionada, este trabajo tiene como objetivo la identificación y

presentación de tópicos más dominantes en trabajos de investigación presentes en el repositorio institucional de la UNAP, mediante el uso de las técnicas como coeficiente angular y nube de palabras. A partir de este análisis de tendencia, serán presentados y discutidos los tópicos con tendencias crecientes, decrecientes, atípicas y neutrales. Como posible impacto esperado a partir de los resultados de este trabajo, se pretende brindar claridad en relación a las tendencias de investigación más relevantes en el repositorio institucional de la UNAP.

El trabajo de investigación está organizado de la siguiente forma: La sección material y métodos presenta una breve descripción de las características de los datos utilizados para este estudio y los abordajes utilizados para el Análisis de tendencia. La sección Resultados y Discusión presentan y discuten los resultados obtenidos a partir del análisis de los datos, el análisis de tendencia y de las nubes de palabras. La sección Conclusiones presenta un resumen de tendencias, las restricciones del abordaje utilizado y las perspectivas de trabajos futuros.

II. MATERIALES Y MÉTODOS

Este estudio analizó los trabajos de investigación (tesis) publicados en el repositorio de la UNAP, este análisis se, que hacen referencia a la totalidad de los trabajos de investigación (pregrado y postgrado) que se encuentran en el repositorio de la UNAP. Los datos utilizados fueron los títulos, año de publicación, área de estudio y escuela profesional. La metodología utilizada fue la siguiente:

1) Extracción y colección de datos del repositorio de la UNAP mediante la técnica de web scraping.

2) Creación de un dataset para los trabajos de investigación, con las variables: Título, Año, Área de estudio y Escuela Profesional.

3) Eliminación de StopWords, tales como los artículos, preposición, pronombres y conjunciones que por su naturaleza tienden a aparecer mucho, lo que generaría que estas palabras sean las que tienen mayor tendencia.

4) Obtención de la frecuencia, que representa cuantitativamente el número de veces que cada palabra aparece en los títulos de un trabajo, como resultado se tiene una lista ordenada del respectivo conteo. Consecuentemente, se tiene los términos más relevantes en los trabajos de investigación publicados en el repositorio de la UNAP.

5) Transformación de la moda a una métrica relativa denominada 'Frecuencia', para ello fue dividida la moda de un determinado año entre la cantidad de trabajos de investigación del mismo año. Esta actividad fue necesaria debido a la diferencia de trabajos publicados por años.

6) Una vez que se generan estos valores relativos por término y año, fue creado el modelo de regresión lineal, a fin de determinar el coeficiente de correlación y el coeficiente angular del modelo, coeficientes de correlación y angular que permita inferir la tendencia en la utilización de un término en base a la frecuencia relativa. Los coeficientes angulares positivos indican una tendencia creciente en la utilización del término (palabra) en los títulos de los trabajos. Los coeficientes negativos, indican una tendencia decreciente.

7) Finalmente, fue usada la técnica de visualización de datos de nube de palabras, que posibilita la comparación

de las frecuencias de palabras (moda). Siendo que las palabras que tienen coeficientes positivos presentan algún tipo de color y tamaño de fuente, mientras que las que

presentan coeficiente negativo presentan un color plumizo y tamaño de fuente.

III. RESULTADOS Y DISCUSIÓN

Con el abordaje propuesto fueron identificados los tópicos de investigación más relevantes en la UNAP. El abordaje tuvo inicio con la extracción y colección de los datos a tratar. De esta manera en la Fig. 1. podemos observar los datos, tanto para pregrado como para postgrado, contando con 6185 y 1961 registros respectivamente, nótese que también fueron extraídos y colectados los datos de Título, Año, Escuela, Área y Link que servirá para el objetivo de nuestro estudio e incluso el desarrollo y divulgación de un sistema informático. En la Fig. 2. Podemos observar los datos tratados, donde fue posible la generación frecuencia a partir de los títulos de trabajo de investigación que se encuentran en el repositorio de la UNAP, observe la palabra Sistema y la cantidad de trabajos, donde evolución de la palabra en el tiempo es creciente a excepción del 2020, quizá se deba a la dinámica social y las actividades académicas debido a la enfermedad del covid-19.

Prueba (Área = Ciencias de la Ingeniería, Escuelas = Todas, Palabra = Sistema)

	Año	Moda	Palabra	Total Trabajos	Frecuencia
0	2003	0	0	1	0.000
1	2004	0	0	1	0.000
2	2005	0	0	0	0.000
3	2006	0	0	4	0.000
4	2007	1	0	2	0.500
5	2008	0	0	4	0.000
6	2009	0	0	15	0.000
7	2010	3	0	39	0.077
8	2011	1	0	24	0.042
9	2012	7	0	39	0.179
10	2013	25	0	96	0.260
11	2014	33	0	120	0.275
12	2015	25	0	155	0.161
13	2016	30	0	172	0.174
14	2017	89	0	453	0.196
15	2018	53	0	359	0.148
16	2019	51	0	380	0.134
17	2020	28	0	134	0.209
18	2021	18	0	83	0.217

escuelas.

Data pregrado, registros totales = 6185

Estructura = Título, Año, Escuela, Área, Link

Título	Año	Escuela	Área	Link
1 Evaluación de la ejecución presu	2016	Escuela Profesional de Ciencias Contables	Ciencias Económicas y Empresarial	http://repositorio.unap.edu.pe/handle/UNAP/1550
2 Analisis de la gestión del riesgo c	2016	Escuela Profesional de Ciencias Contables	Ciencias Económicas y Empresarial	http://repositorio.unap.edu.pe/handle/UNAP/1551
3 Evaluación de la gestión del proc	2016	Escuela Profesional de Administración	Ciencias Económicas y Empresarial	http://repositorio.unap.edu.pe/handle/UNAP/1552
4 Factores que influyen en la actua	2016	Escuela Profesional de Ingeniería Estadística	Ciencias de la Ingeniería	http://repositorio.unap.edu.pe/handle/UNAP/1553
5 Proceso de Selección de Persona	2016	Escuela Profesional de Administración	Ciencias Económicas y Empresarial	http://repositorio.unap.edu.pe/handle/UNAP/1555
6 Evaluación comparativa de las Me	2016	Escuela Profesional de Administración	Ciencias Económicas y Empresarial	http://repositorio.unap.edu.pe/handle/UNAP/1557
7 Analisis comparativo del derecho	2016	Escuela Profesional de Derecho	Ciencias Sociales	http://repositorio.unap.edu.pe/handle/UNAP/1569
8 Modelamiento del Impacto Ambi	2015	Escuela Profesional de Ingeniería de Sistem	Ciencias de la Ingeniería	http://repositorio.unap.edu.pe/handle/UNAP/1669
9 DATAMART de Contrataciones Pú	2014	Escuela Profesional de Ingeniería de Sistem	Ciencias de la Ingeniería	http://repositorio.unap.edu.pe/handle/UNAP/1670
10 Conocimiento de los Adolescent	2014	Escuela Profesional de Enfermería	Ciencias Biomédicas	http://repositorio.unap.edu.pe/handle/UNAP/1671
11 El Grado de desarrollo de la Histi	2013	Escuela Profesional de Educación Física	Ciencias Sociales	http://repositorio.unap.edu.pe/handle/UNAP/1672
12 Percepciones Sociales de los Age	2013	Escuela Profesional de Sociología	Ciencias Sociales	http://repositorio.unap.edu.pe/handle/UNAP/1673
13 Estudio comparativo de los fact	2013	Escuela Profesional de Medicina Humana	Ciencias Biomédicas	http://repositorio.unap.edu.pe/handle/UNAP/1674
14 Aplicación web de bolsa de empl	2014	Escuela Profesional de Ingeniería de Sistem	Ciencias de la Ingeniería	http://repositorio.unap.edu.pe/handle/UNAP/1675
15 Sistema de información para el r	2015	Escuela Profesional de Ingeniería Estadística	Ciencias de la Ingeniería	http://repositorio.unap.edu.pe/handle/UNAP/1676
16 Sistema Eip Aplicando Scrum Par	2015	Escuela Profesional de Ingeniería de Sistem	Ciencias de la Ingeniería	http://repositorio.unap.edu.pe/handle/UNAP/1677
17 Factores Asociados a la capacida	2014	Escuela Profesional de Enfermería	Ciencias Biomédicas	http://repositorio.unap.edu.pe/handle/UNAP/1678
18 Caracterización de redes Minen	2014	Escuela Profesional de Ingeniería Metalúrgica	Ciencias de la Ingeniería	http://repositorio.unap.edu.pe/handle/UNAP/1679
20 Relación entre Patrón Alimentari	2013	Escuela Profesional de Nutrición Humana	Ciencias Biomédicas	http://repositorio.unap.edu.pe/handle/UNAP/1680
21 Evaluación del Modelo "QUARTE	2014	Escuela Profesional de Ingeniería Civil	Ciencias de la Ingeniería	http://repositorio.unap.edu.pe/handle/UNAP/1681
22 Analisis de la Sombra y el sombri	2014	Escuela Profesional de Ingeniería de Sistem	Ciencias de la Ingeniería	http://repositorio.unap.edu.pe/handle/UNAP/1682
23 Diseño de una Red de Banda Anc	2014	Escuela Profesional de Ingeniería Electrónica	Ciencias de la Ingeniería	http://repositorio.unap.edu.pe/handle/UNAP/1683

Data postgrado, registros totales = 1961

Estructura = Título, Año, Escuela, Área, Link

Título	Año	Escuela	Área	Link
1 Influencia del material educativo elaborado con	2011	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/185
2 Importancia del juego "el zorro y la oveja" para	2011	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/186
3 Revalorando la lectura de señas y señaleros en	2010	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/187
4 La producción de cuentos andinos a partir de im	2014	Maestría en Lingüística Andina y 2. MAESTRIA		http://repositorio.unap.edu.pe/handle/UNAP/188
5 Limitaciones en el aprendizaje del castellano en	2010	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/189
6 Cuentos andinos para la producción de textos en	2010	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/190
7 Revalorar la tecnología andina en la confección	2010	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/191
8 El programa currculo de mejoramiento de la calid	2010	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/192
9 El material audiovisual en relación a la oralidad	2011	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/193
10 Uso del abecedario andino en el aprendizaje de la	2010	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/194
11 La mitología andina para la formación de valores	2011	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/195
12 Participación comunitaria en la aplicación del pr	2010	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/196
13 Uso de estrategias metodológicas en eib en los	2010	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/197
14 Aprendizaje de Aymara como segunda lengua por	2007	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/198
15 Influencia de las señas y señaleros en el aprendi	2012	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/199
17 Enseñanza y aprendizaje del castellano como seg	2010	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/200
18 Importancia de la lengua originaria en el aprendi	2011	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/201
19 Estrategias metodológicas para desarrollar habi	2014	Maestría en Lingüística Andina y 2. MAESTRIA		http://repositorio.unap.edu.pe/handle/UNAP/202
20 La onomatopeya como estrategia didáctica para	2013	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/203
21 La autonomía como desarrollo fundamental en l	2013	Segunda Especialización en Edu.3. SEGUNDA ESPECIALIZACIÓN		http://repositorio.unap.edu.pe/handle/UNAP/204

Fig. 1. Datos de pregrado y postgrado de la UNAP.

En la Fig. 3. Es presentado ejemplos con los coeficientes de correlación y angular, dos métricas que servirán para analizar la tendencia de una determinada palabra de interés. Así por ejemplo tenemos las palabras **crecientes**: Sistema (0.43; 0.01) y Región (0.73; 0.0017), las palabras **decrecientes**: Puno (-0.55; -0.0009) y Juliaca (-0.29; -0.0007), las palabras **intermedias**: Empresa (0.06; 0.0009) y Administración (0.06; 0.0009) y las palabras **atípicas**: Leyenda (0.0; 0.0) y Sociedad (0.0; 0.0).

Crecientes:

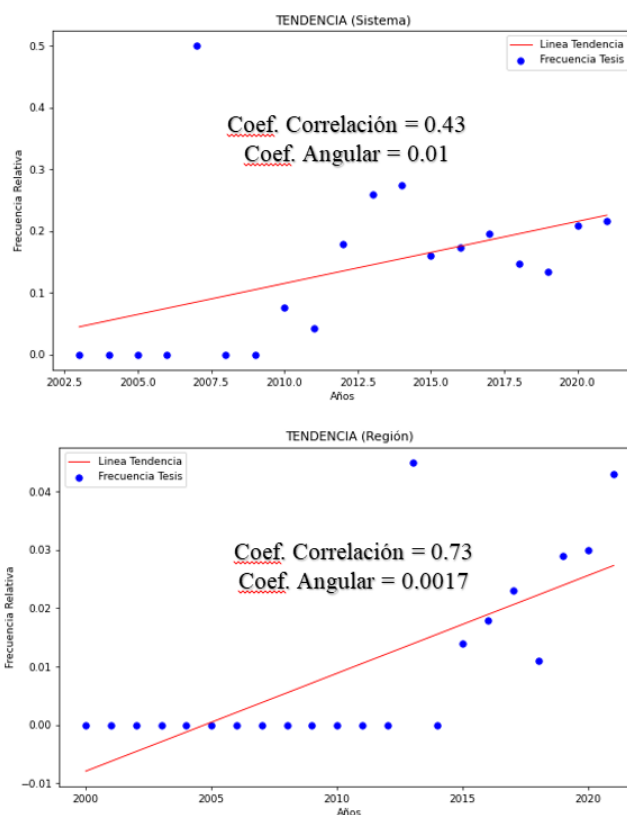
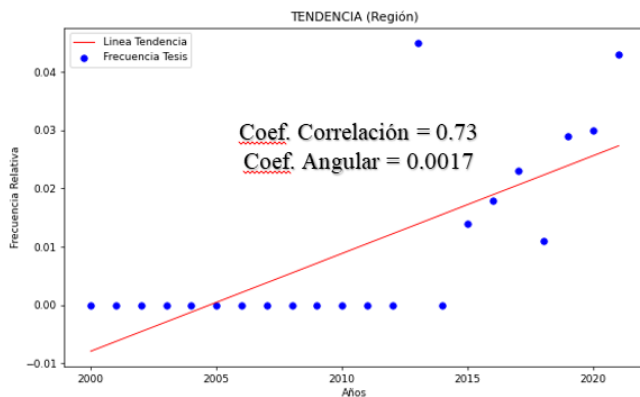
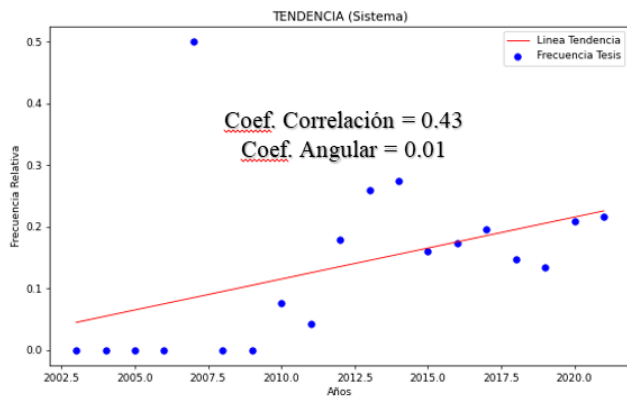
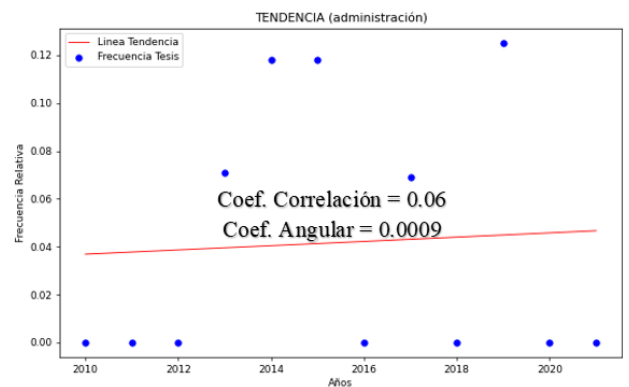
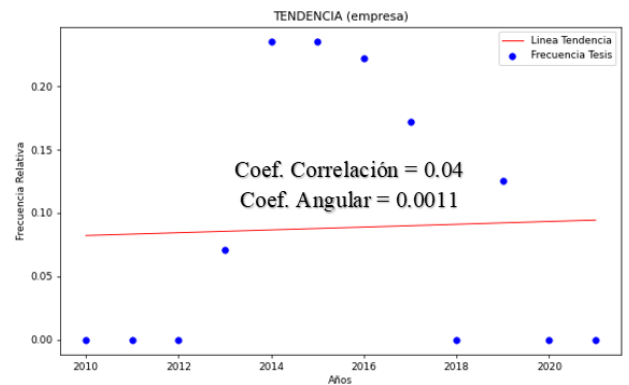


Fig. 3. Gráficos de las palabras con mayor tendencia (Crecientes), con menor tendencia (decreciente), con tendencia neutral (Intermedio) y aquellos que son atípicos por su escasa presencia en el tiempo.

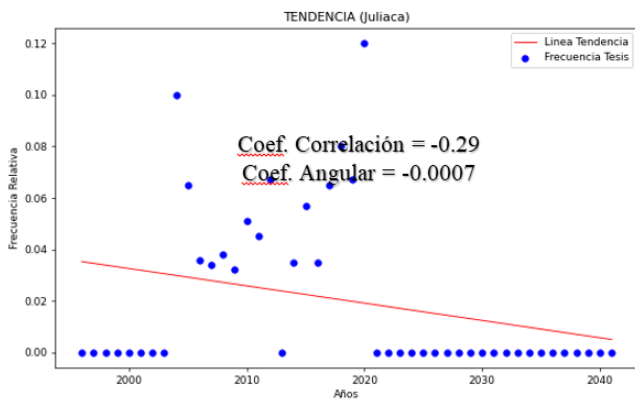
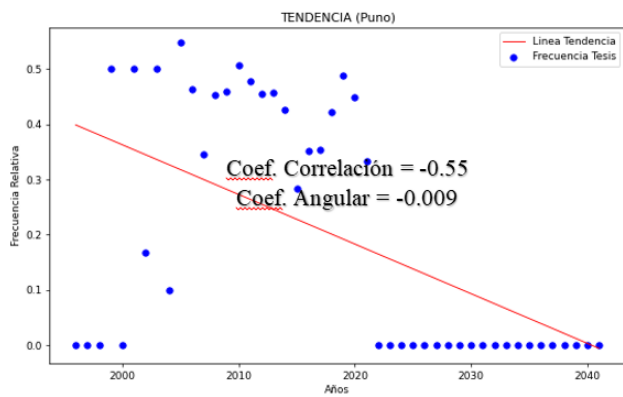
Crecientes:



Intermedios:



Decrecientes:



Atípico:

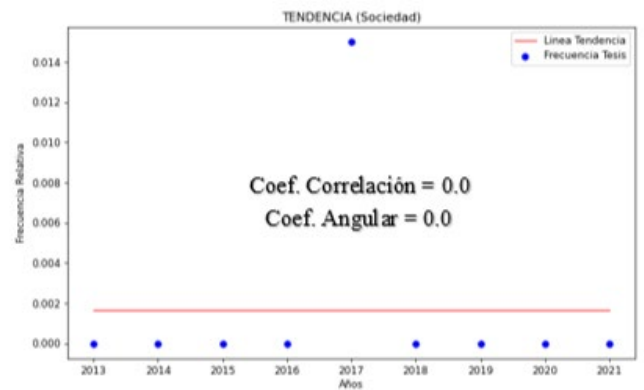
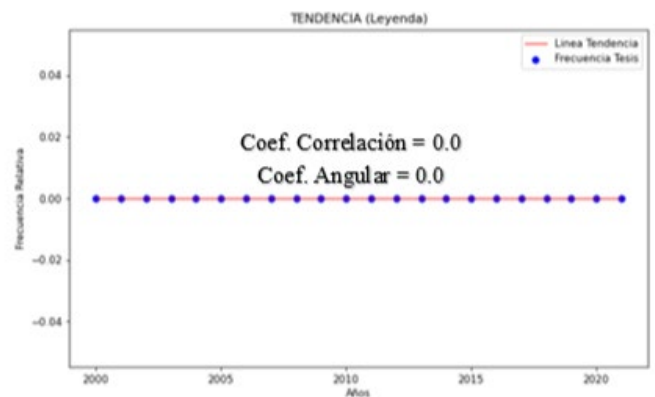


Fig. 3. Gráficos de las palabras con mayor tendencia (Crecientes), con menor tendencia (decreciente), con tendencia neutral (Intermedio) y aquellos que son atípicos por su escasa presencia en el tiempo.

IV. CONCLUSIÓN

Este trabajo tuvo como objetivo analizar las tendencias de investigación de la UNAP a partir de los títulos de los trabajos publicados que se encuentran en el repositorio de la UNAP. Fueron identificadas las palabras con mayor frecuencia en los títulos y posteriormente, a partir de las palabras más frecuentes de cada año, fue realizada la ponderación por la cantidad de los trabajos realizados en un año, obteniéndose la frecuencia relativa de las palabras. Fue utilizada el modelo de regresión a fin de determinar el coeficiente de correlación y coeficiente angular para el uso de las palabras de interés en el repositorio de la UNAP.

V. BIBIOGRAFÍA

[1] Ahearn, L. M. (2014). Detecting research patterns and paratextual features in AE word clouds, keywords, and titles. *American ethnologist*, 41(1), 17-30.

[2] Ahmed, Z. (2017). Data management and big data text analytics. In *Special Issue-National Conference on" Novel Trends in Computer Science"*(TECHSA-17) p (pp. 140-144).

[3] Benckendorff, P., & Zehrer, A. (2013). A network analysis of tourism research. *Annals of Tourism Research*, 43, 121-149.

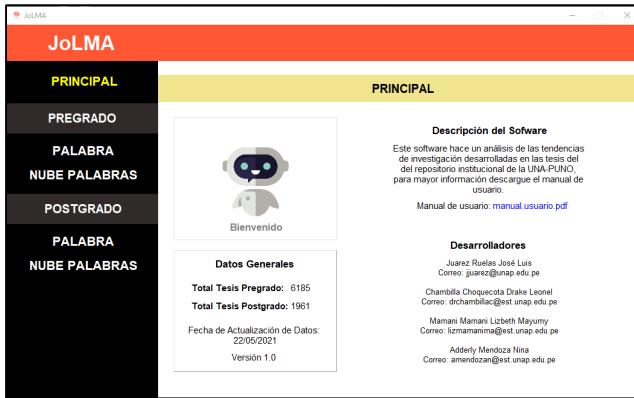
[4] Diouf, R., Sarr, E. N., Sall, O., Birregah, B., Bousso, M., & Mbaye, S. N. (2019, December). Web Scraping: State-of-the-Art and Areas of Application. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 6040-6042). IEEE.

[5] Haugerud, A. (2013). Editor's foreword: AE's keywords by decade. *American Ethnologist*, 40(1), 1-5.

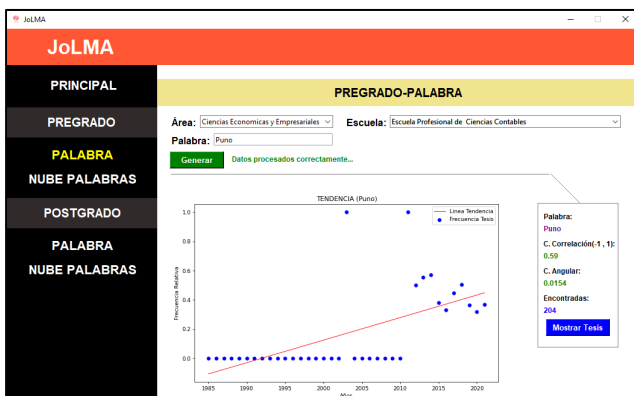
[6] Hunt, C. A., Gao, J., & Xue, L. (2014). A visual analysis of trends in the titles and keywords of top-ranked tourism journals. *Current Issues in Tourism*, 17(10), 849-855.

[7] Ryan, C. (2005). The ranking and rating of academics and journals in tourism research. *Tourism Management*, 26(5), 657-662.

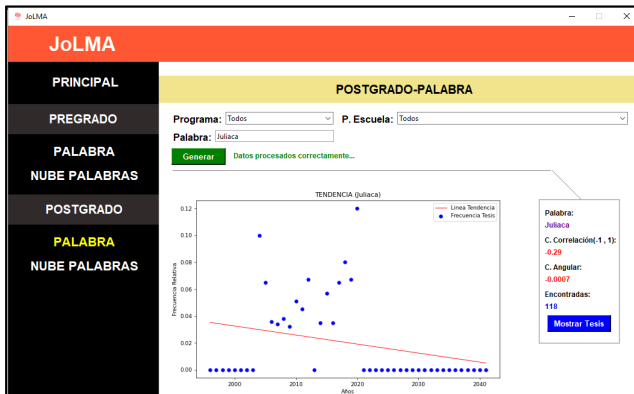
Principal:



Pregrado Palabra:



Pregrado Palabra:



Pregrado Palabra:



Fig. 4. Gráficos del sistema de información, análisis de tendencias y nube de palabras a nivel de pregrado y postgrado.