



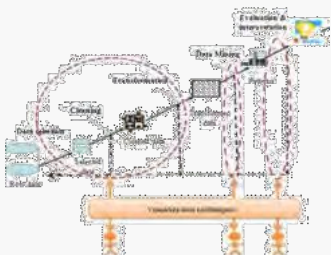
VISUALIZANDO LA INFORMACIÓN EN CIENCIA DE DATOS

María Alejandra Malberti Graciela E. Beguerí Raúl Oscar Klenzi Manuel Ortega Luis Olguín Fabrizio Amaya Joaquín Cortez
 amalberti@gmail.com grabeda@gmail.com rauloscarklenzi@gmail.com manuel.ortega@gmail.com lolguinunsj@gmail.com fabrizio.amaya88@gmail.com joaquinortez1999@gmail.com

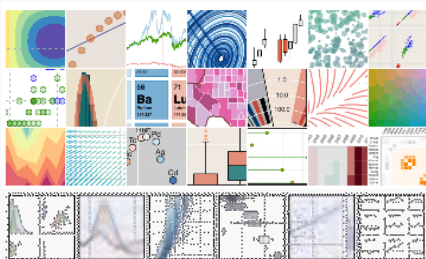
Departamento de Informática (DI)—Instituto de Informática (idei)
 Facultad de Ciencias Exactas, Físicas y Naturales—Universidad Nacional de San Juan

CONTEXTO

El Proyecto articula líneas de investigación de un grupo de investigadores y docentes de las carreras Licenciatura en Sistemas de Información y Licenciatura en Ciencias de la Computación del Departamento de Informática de la Facultad de Ciencias Exactas Físicas y Naturales de la Universidad Nacional de San Juan (FCEFN-UNSJ), y se encuentra contenido en el Laboratorio de Sistemas Inteligentes para Extracción de Conocimiento en Datos Masivos del Instituto de Informática de la misma facultad. Asimismo, el proyecto se encuentra aprobado y subsidiado por Consejo de Investigaciones Científicas y Técnicas y de Creación Artística CICITCA- UNSJ.



Integración de técnicas de visualización. Mohamed, E.



Librerías Bokeh y Seaborn <https://www.iartificial.net/librerias-de-python-para-machine-learning>

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las tareas que se están llevando a cabo comprenden varias líneas de investigación relacionadas con:

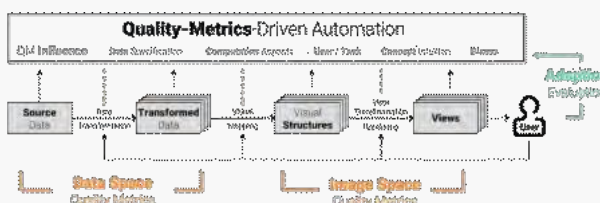
- Ciencia de Datos, principalmente lo relativo a Visualización de Información.
- Deep Learning a través del análisis y diferencias entre Boosted Decision Tree Regression y Redes Neuronales del tipo LSTM (Long Short Term Memory) en entorno de visualización open source Knowledge.
- Herramientas de software libre para arquitecturas secuenciales, paralelas y distribuidas particularmente KNIME Analytics cuyo enorme potencial didáctico permite visualizar mediante workflows las diferentes etapas que constituyen el pipeline del Data Science.
- Lenguajes de programación de código abierto tales como Python, JavaScript y R.

FORMACIÓN DE RECURSOS HUMANOS

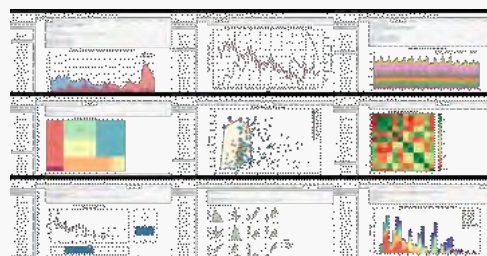
Los alumnos de grado se hallan realizando sus trabajos finales en las líneas de investigación mencionadas.

Miembros del equipo dirigen y asesoran trabajos finales de grado y posgrado; así como becas de Investigación, categoría Alumnos Avanzados CICITCA y CIN.

Como aporte a la sociedad se ha planificado un curso destinado a las pequeñas y medianas empresas de la provincia de San Juan, a partir de una iniciativa del gobierno y con el aval éste. Dicho curso tiene como finalidad introducir los conceptos básicos de Ciencia de Datos y mostrar las virtudes o ventajas de extraer conocimiento de los datos.



Análisis Visual basada en métricas de calidad con la capa adicional de automatización (Behrisch, et al., 2018)



Cuadernos de Jupyter Notebook integrados en Azure Cosmos DB <https://docs.microsoft.com/es-es/azure/cosmos-db/cosmosdb-jupyter-notebooks>

RESULTADOS

Se está realizando, sobre diversos conjuntos de datos abiertos, una comparativa entre las bibliotecas de visualización disponibles en el lenguaje Python. Hasta ahora las bibliotecas exploradas son: Matplotlib, Seaborn, Plotly, Bokeh, Altair, y Folium.

También, se han creado varios Jupyter notebooks, con diversos conjuntos de datos libres, con los que se ejemplifican el uso de las bibliotecas.

En el caso particular del espacio vectorial de las palabras, área del procesamiento del lenguaje natural, se aplicaron técnicas de aprendizaje automático basadas en redes neuronales y se empleó el método Word2vec mediante la librería Gensim de Python, para su visualización.

Está en desarrollo una tesis de maestría con datos temporales,

WICC 2021 – XXIII Workshop de Investigadores en Ciencias de la Computación



